Cognition Assessment in Virtual Reality (CAVIR)

Associations with neuropsychological performance and activities of daily living in patients with mood or psychosis spectrum disorders

Jespersen, Andreas E.; Lumbye, Anders; Schandorff, Johanna; Damgaard, Viktoria; Glenthøj, Louise B.; Nordentoft, Merete; Mikkelsen, Christina; Didriksen, Maria; Ostrowski, Sisse R.; Vinberg, Maj; Wæhrens, Eva E.; Miskowiak, Kamilla W.

Published in:
Journal of Affective Disorders

DOI:
10.1016/j.jad.2024.10.095

Publication date:
2025

Document version
Publisher's PDF, also known as Version of record

Journal of Affective Disorders 369 (2025) 1053–1063

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Cognition Assessment in Virtual Reality (CAVIR): Associations with neuropsychological performance and activities of daily living in patients with mood or psychosis spectrum disorders

Andreas E. Jespersen [a,b], Anders Lumbye [c], Johanna Schandorff [a,b], Viktoria Damgaard [a,b], Louise B. Glenthøj [b,d], Merete Nordentoft [d,e], Christina Mikkelsen [f,g], Maria Didriksen [f], Sisse R. Ostrowski [e,f], Maj Vinberg [e,h], Eva E. Wæhrens [i,j], Kamilla W. Miskowiak [a,b,*,1,2]

[a] Neurocognition and Emotion in Affective Disorders (NEAD) Centre, Psychiatric Centre Copenhagen, Frederiksberg Hospital, Copenhagen, Denmark
[b] Department of Psychology, University of Copenhagen, Denmark
[c] Wide Angle Media, Copenhagen, Denmark
[d] Copenhagen Research Centre for Mental Health-CORE, Mental Health Centre Copenhagen, Copenhagen University Hospital, Denmark
[e] Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
[f] Department of Clinical Immunology, Copenhagen University Hospital - Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark
[g] Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[h] The Early Multimodular Prevention and Intervention Research Institution (EMPIRI), Mental Health Centre, Northern Zealand, Mental Health Services, Denmark
[i] The Parker Institute, Bispebjerg and Frederiksberg Hospital, University of Copenhagen, Copenhagen, Denmark
[j] Occupational Science, User Perspectives and Community-based Interventions, Department of Public Health, University of Southern Denmark, Odense, Denmark

ARTICLE INFO

Keywords:
Virtual reality
Activities of daily living
Cognition
Depression
Bipolar disorder
Schizophrenia

ABSTRACT

Background: More ecologically valid tools are needed to better capture daily-life cognitive impairments in patients with mood or psychosis spectrum disorders in clinical settings and cognitive treatment trials. We developed the Cognition Assessment in Virtual Reality (CAVIR) test, which assesses daily-life cognitive skills in an immersive virtual reality kitchen scenario. This study investigated the validity and sensitivity of CAVIR, including its association with activities of daily living (ADL) ability.

Methods: Seventy symptomatically stable patients with mood or psychosis spectrum disorders and 70 healthy controls completed CAVIR and standard neuropsychological tests and were rated for clinical symptoms, functional capacity, and subjective cognition. In addition, patients' ADL ability was evaluated with the Assessment of Motor and Process Skills.

Results: Higher global CAVIR performance correlated moderately with better global neuropsychological test scores ($r_s(138) = 0.60$, $p < 0.001$) and showed a weak to moderate association with better ADL process ability in patients ($r(45) = 0.40$, $p < 0.01$), also after adjusting for sex and age ($p_s \leq 0.03$). In comparison, neuropsychological performance, interviewer- and performance-based functional capacity, and subjective cognition were not significantly associated with ADL process ability ($p_s \geq 0.09$). Further, CAVIR was sensitive to cognitive impairments in patients and was able to differentiate between patients with and without the ability to undertake regular employment.

Limitations: The modest sample size and concomitant medication.

* Corresponding author at: NEAD Centre, Psychiatric Centre Copenhagen, Frederiksberg Hospital, Hovedvejen 17, DK-2000 Copenhagen, Denmark.
E-mail addresses: Andreas.elleby.jespersen@regionh.dk (A.E. Jespersen), johanna.mariegaard.schandorff@regionh.dk (J. Schandorff), viktoria.damgaard.01@regionh.dk (V. Damgaard), Louise.birkedal.glenthoej@regionh.dk (L.B. Glenthøj), Merete.nordentoft@regionh.dk (M. Nordentoft), Christina.mikkelsen@regionh.dk (C. Mikkelsen), Maria.didriksen@regionh.dk (M. Didriksen), Sisse.Rye.Ostrowski@regionh.dk (S.R. Ostrowski), Maj.vinberg@regionh.dk (M. Vinberg), Eva.elisabet.waehrens@regionh.dk (E.E. Wæhrens), kamilla.woznica.miskowiak@region.dk, Kamilla.woznica.miskowiak@regionh.dk (K.W. Miskowiak).
[1] **Location of work and address for reprints**: Neurocognition and Emotion in Affective Disorders (NEAD) Centre, Psychiatric Centre Copenhagen, Frederiksberg Hospital, Hovedvejen 17, DK-2000 Copenhagen, Email: Kamilla.woznica.miskowiak@regionh.dk
[2] **Host institution:** Psychiatric Center Copenhagen, Mental Health Services in the Capital Region of Denmark, Kristineberg 3, DK-2100, Copenhagen, + (45) 38640000

https://doi.org/10.1016/j.jad.2024.10.095
Received 19 June 2024; Received in revised form 27 September 2024; Accepted 20 October 2024
Available online 22 October 2024

Conclusion: Our results indicate that CAVIR is a sensitive measure of daily-life cognitive skills in patients with mood or psychosis spectrum disorders.

1. Introduction

Persistent cognitive impairments are common in both mood disorders (MD) and psychosis spectrum disorders (PD) (Bora and Pantelis, 2015; Bourne et al., 2013; Green et al., 2020; Rock et al., 2014), significantly affecting daily-life functioning, quality of life, and workforce capacity (Green et al., 2000; Jaeger et al., 2006; Jaeger and Vieta, 2007; Tse et al., 2014). Consequently, improving cognitive functioning has become a key treatment target for enhancing functional recovery (Bowie et al., 2006; Miskowiak et al., 2018; Tsapekos et al., 2020). A recent systematic review by the International Society for Bipolar Disorder (ISBD) Targeting Cognition Task Force found preliminary pro-cognitive effects of several pharmacological and psychological interventions (Miskowiak et al., 2022a). Among these, modafinil and lurasidone showed preliminary positive results, whereas cognitive remediation therapy (CRT) showed the most consistent cognitive benefits. Indeed, meta-analytic evidence has shown small-to-moderate effects of CRT on neuropsychological test performance across MD and PD (Goldberg et al., 2023; Samamé et al., 2023; Vita et al., 2021). However, there remains a significant gap between these test improvements and real-world functioning (Goldberg et al., 2023; Miskowiak et al., 2022a; Samamé et al., 2023). This discrepancy highlights the need for novel cognitive outcome measures in treatment trials that better reflect daily-life cognitive skills.

A key methodological challenge is accurately capturing patients' cognitive skills required for daily tasks (Miskowiak et al., 2017; Miskowiak et al., 2018). Neuropsychological tests – the primary outcomes in cognition trials – have *limited ecological validity* as they are administered in a quiet, controlled environment with specific instructions, which bears little resemblance to patients' real-life cognitive hassles (Miskowiak et al., 2017). Accordingly, neuropsychological performance accounts for only 5–21 % of the variance in daily-life functioning, which limits the insight into whether treatment-related cognitive gains transfer into improved functional outcomes (Van der Elst et al., 2008). The ISBD therefore recommends assessing transfer effects to daily-life functioning as a secondary outcome (Miskowiak et al., 2022a) using the *interview-based* Functional Assessment Short Test (FAST) (Rosa et al., 2007) or the *performance-based* University of California, San Diego Performance-based Skills Assessment - Brief (UPSA-B) (Mausbach et al., 2007). However, the FAST includes measures like income status that do not necessarily improve with pro-cognitive interventions (Rosa et al., 2007), and self-report measures are based on patients' perceptions and can be influenced by symptom severity (Ott et al., 2019; Petersen et al., 2019). Performance-based measures like the UPSA-B are more closely associated with objective neuropsychological performance but are marked by ceiling effects in psychiatric disorders (Ott et al., 2019; Østergaard Christensen et al., 2014). These limitations underscore the need for more ecologically valid cognitive measures that accurately reflect real-world functioning without being influenced by patients' insight or symptom severity (Miskowiak et al., 2017).

Virtual reality (VR) platforms may offer a potential solution by simulating complex real-life like cognitive challenges in controlled environments (Bohil et al., 2011; Neguț et al., 2016; Parsons, 2015). While several desktop VR tools have been validated for cognitive- or functional capacity assessment in psychiatric and neurological populations (Hørlyck et al., 2021; Keefe et al., 2016; Rand et al., 2009), their ecological validity is limited compared to immersive VR devices like head-mounted displays (HMDs). Studies indicate that greater immersion enhances the feeling of *presence* and, thereby, user engagement (Makransky and Petersen, 2021; Makransky et al., 2019), which may improve the ecological validity of cognitive assessment. This has led to interest in designing and testing immersive VR tools for cognitive assessment in psychiatric populations, including MD and PD (Adriasola et al., 2024; Mannan et al., 2023; Voinescu et al., 2023).

We recently developed an immersive HMD VR cognition test, the Cognition Assessment in Virtual Reality (CAVIR), in which participants complete five tasks related to planning and preparing a meal in a kitchen scenario (Miskowiak et al., 2022b). In line with WHO's International Classification of Functioning, Disability and Health (ICF) (Comment WHO, 2013), the CAVIR test was designed to assess *cognitive skills* necessary for tackling typical daily-life challenges (i.e., ICF level of activity/participation). The initial version of CAVIR showed good feasibility and a moderate association with traditional neuropsychological tests in patients with MD or PD (Miskowiak et al., 2022b). Key limitations were, however, that some subtasks were marked by ceiling effects and that the test was only validated against neuropsychological tests (Miskowiak et al., 2022b) rather than real-world functional capacity measures capturing activities of daily living (ADLs) (Kirkham et al., 2024). Indeed, for a cognitive test to have ecological validity, it must predict performance in similar real-world situations, termed *veridicality* (Franzen and Wilhelm, 1996; Spooner and Pachana, 2006). An ADL ability measure such as the Assessment of Motor and Process Skills (AMPS) (Fisher and K. B. J., 2012) overcomes the limitations of the FAST and UPSA-B by directly assessing functioning in real-world settings like a standardised testing apartment through tasks with individually adjusted difficulty levels. Notably, the AMPS can assess ADL ability specifically related to household tasks such as meal preparation (Fisher and K. B. J., 2012) and has been shown to be an important predictor of occupational capacity (Haslam et al., 2010; Vandamme, 2010). There is thus a need to explore the associations between CAVIR, neuropsychological performance and AMPS to elucidate whether the test can bridge the gap between standard cognitive tests and real-world functioning to better capture daily-life cognitive skills.

This study therefore aimed to investigate the convergent validity and sensitivity of the optimised CAVIR in a sample of symptomatically stable patients with MD or PD and healthy controls (HC). We hypothesised (i) that global CAVIR scores would be positively associated with global performance on traditional neuropsychological tests across patients and HCs, (ii) that global CAVIR performance would be positively associated with patients' real-life ADL ability evaluated with the AMPS, and (iii) that patients would display impairment on the global CAVIR score compared to HCs. For exploratory purposes, we investigated the validity and sensitivity of each CAVIR subtask and the associations between CAVIR and FAST, UPSA-B and subjective cognition. Finally, we explored the ability of CAVIR to differentiate between patients according to their occupational capacity based on ADL ability.

2. Methods

2.1. Participants and recruitment

Data was collected from the baseline assessments of two ongoing clinical trials at the University of Copenhagen, Denmark: a trial investigating the effect of VR-based cognitive remediation (Clinicalrials.gov Identifier: NCT06038955) and a trial investigating hypoxia cognition training (Clinicalrials.gov Identifier: NCT06121206). Combining baseline data from these trials was deemed appropriate due to their shared participant inclusion criteria and cognitive, functioning, and symptom assessments, as described in more detail in the study protocol (Jespersen et al., 2024a). In brief, eligible participants were 18–55 years of age and

fluent in Danish. All patient participants were recruited from outpatient clinics in the Capital Region of Denmark. Patients with MD had an ICD-10 diagnosis of either unipolar disorder ($n = 16$) or bipolar disorder ($n = 36$) and were in full or partial remission, as reflected by scores ≤14 on the Hamilton Depression Rating Scale 17 items (HRDS-17) (Hamilton, 1960) and the Young Mania Rating Scale (YMRS) (Young et al., 1978). Patients with PD had an ICD-10 diagnosis within the F20-spectrum of schizophrenia ($n = 7$), schizotypal disorder ($n = 9$), or unspecified non-organic psychosis ($n = 2$) and were all assessed to be relatively symptom-stable by their treating clinician upon referral. Diagnoses were confirmed using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interview (Wing et al., 1990). The HC participants were recruited through website advertisements ($n = 37$) or from blood banks in the Capital Region of Denmark ($n = 33$) and were free of any personal history of psychiatric illness confirmed using the SCAN interview. General exclusion criteria were current substance use, neurological disorder, severe somatic illness, dyslexia, a daily use of benzodiazepines >22,5 mg oxazepam, and electroconvulsive therapy (ECT) within the past three months. Both trials were approved by the local ethics committee in the Capital Region of Denmark. The recommendations of the Declaration of Helsinki were followed, and written informed consent was obtained from all participants prior to study participation.

2.2. Procedure

The assessments were conducted over two days. On day 1, all participants were assessed with CAVIR and a traditional neuropsychological test battery (details below). Participants were also assessed on functional capacity, completed questionnaires concerning subjective cognitive complaints, and underwent mood ratings with the HDRS-17 (Hamilton, 1960) and YMRS (Young et al., 1978). For participants with PD, positive symptoms were assessed using the Assessment of Positive Symptoms (SAPS) (Andreasen, 1984) and negative symptoms using the Brief Negative Symptom Scale (BNSS) (Kirkpatrick et al., 2011). On day 2, a sub-sample of the patient participants underwent an evaluation of their ADL ability with the AMPS (Fisher and K. B. J., 2012), which was only part of the assessment in the VR cognitive remediation trial (Jespersen et al., 2024a). The AMPS evaluation was arranged within seven days of the first visit and with as few days apart as possible. The assessments on day 1 and day 2 were conducted independently, and the assessors were blinded to the outcome of the other assessment.

Fig. 1. 1A: The 360° CAVIR kitchen environment administered on a Meta Quest 128 GB portable headset. 1B: Task 1: memorising and finding ingredients from a list (verbal memory); 1C: Task 2: planning the order in which to complete subtasks involved in meal preparation (executive function); 1D: Task 3: placing as many correct ingredients as possible in a pot within a time frame (processing speed); 1E: Task 4: memorising the location of flatware and kitchen utensils (working memory); 1F: Task 5: repeatedly checking the food in the oven in response to specific cues so the food does not get burned (attention). Participants in the current study completed the Danish version of CAVIR, but for the reader's sake, the pictures included here are from the English version.

2.3. Materials

2.3.1. Cognition Assessment in Virtual Reality (CAVIR)

CAVIR is an immersive, self-administered 360° VR test in a kitchen scenario, where the participant is assessed on daily-life cognitive tasks related to preparing a meal (Miskowiak et al., 2022b) (see Fig. 1 and Table 1). The test lasts 15 min and is administered on a standalone head-mounted Meta Quest 128 GB portable headset. Through the headset, participants are instructed by a pre-recorded voice to carry out five subtasks probing different cognitive skills and underlying cognitive functions: (1) memorising ingredients from a list (verbal memory; Fig. 1B), (2) planning the order in which to complete subtasks involved in meal preparation (executive functions; Fig. 1C), (3) placing as many correct ingredients as possible in a pot within a time frame (processing speed; Fig. 1D), (4) memorising the location kitchen utensils and flatware (working memory; Fig. 1E) and (5) repeatedly checking the food in the oven in response to specific cues (attention; Fig. 1F). An initial CAVIR version showed a moderate association with standard neuropsychological tests, high levels of presence and low levels of simulation sickness in patients with MD and PD (Miskowiak et al., 2022b). However, subtasks 1, 2, and 5 showed ceiling effects and were, therefore, slightly optimised in the current version of CAVIR, which involved creating a higher difficulty level. Additionally, we developed a parallel version of the optimised CAVIR involving the same five subtasks with different test stimuli and placement of items. All participants in the current study were tested with the optimised version of CAVIR (Danish), with each half receiving one of the two parallel versions in a randomised design. See Appendix A for details on CAVIR technical setup, subtasks and optimisations to the initial version.

Table 1
Overview of the Cognition Assessment in Virtual Reality (CAVIR) subtask measures used in the current study and associated traditional neuropsychological tests assessing verbal learning, executive function, processing speed, working memory, and attention.

Cognitive domains	CAVIR subtask measures	Neuropsychological tests
Verbal learning and memory	Task 1: *d'* score: the ability to discriminate between correct ingredients remembered from the list and false positives (i.e., wrong ingredients) in the fridge/cupboard	RAVLT subtests (IV total, Immediate recall, delayed recall, recognition)
Executive function	Task 2: Number of cooking tasks correctly placed on a to-do list to ensure timely completion (score range: 0–11)	Trail Making B One-Touch Stockings of Cambridge mean choices to correct (CANTAB) Fluency test (S and D) Wisconsin Card Sorting Test
Processing speed	Task 3: Number of correct ingredients placed in the stove pot within 90 s (no score range)	RBANS coding test Trail Making A
Working memory	Task 4: Number of drawers opened until all correct cutlery and flatware have been found (scores are inversed as lower scores indicate better performance)	WAIS Letter-number sequencing Spatial working memory error (CANTAB) Spatial working memory strategy (CANTAB)
Attention	Task 5: *d'* score: The ability to discriminate between correct stimuli (i.e., opening the stove door or regulating temperature in response to correct visual and auditory cues) and false positives (i.e., opening the stove door or regulating temperature in response to foil stimuli)	Rapid visual processing accuracy (CANTAB) Rapid visual processing latency (CANTAB) RBANS digit span

2.3.2. Neuropsychological assessment

Neuropsychological functioning was assessed with a broad test battery comprising the following traditional neuropsychological tests: the One Touch Stocking of Cambridge (OTS), the Spatial Working Memory test (SWM) and the Rapid Visual Information Processing (RVP) from CANTAB (Cambridge Cognition Ltd.), the Rey Auditory Verbal Learning Test (RAVLT) (Schmidt, 1996), WAIS-III letter-number sequencing (Wechsler, 1997), RBANS Coding and Digit span (Randolph et al., 1998), verbal fluency ('d' and 's') (Borkowski et al., 1967), and Trail Making Test A and B (Army Individual Test Battery, 1944). Premorbid verbal IQ was assessed with the Danish Adult Reading Test adapted from the National Adult Reading Test (Crawford et al., 1987).

2.3.3. Assessment of Motor and Process Skills (AMPS)

The AMPS is a standardised occupational therapy evaluation of a person's abilities to perform ADL tasks (Fisher and K. B. J., 2012), which has proven valid in psychiatric populations (Ayres and John, 2015; Pan and Fisher, 1994; Träger et al., 2017). In the current study, patients were assessed on ADL ability related to meal preparation and other household tasks in a standardised test apartment (see Fig. A1 in Appendix A). When administering the AMPS, the participant first undergoes an interview regarding their usual ADL task performance to identify standardised tasks that are relevant and of sufficient challenge. The participant then chooses and performs minimum two ADL tasks under the observation of an AMPS-calibrated occupational therapist. The therapist rates the quality of the participant's ADL performance on 16 ADL *motor* skills (physical effort) and 20 ADL *process* skills (efficiency) based on ease, efficiency, safety and independence. This study focused specifically on ADL *process* skills, as these seem to be most closely related to cognitive impairment in psychiatric disorders (Träger et al., 2017). The AMPS is useful for determining occupational capacity, with an ADL motor score < 2.5 logits and a process score < 1.2 logits indicating insufficient ability to undertake regular employment (Vandamme, 2010).

2.3.4. Assessment of functioning

Functioning was also assessed with the FAST, an *interviewer-based* rating scale (Rosa et al., 2007). Participants in the VR cognitive remediation trial ($n = 33$ HCs and $n = 47$ patients) (Jespersen et al., 2024a) also completed the UPSA-B, a *performance-based* measure of functioning that assesses financial and communication skills (Mausbach et al., 2007; Mausbach et al., 2010).

2.3.5. Assessment of subjective cognition

Participants completed the Cognitive Complaints in Bipolar Disorder Rating Assessment (COBRA), a 16-item questionnaire assessing subjective cognitive difficulties in daily life situations (Rosa et al., 2013). Participants in the VR cognitive remediation trial also completed the Cognitive Difficulties in Everyday Life (CODEL), a similar 16-item questionnaire developed in-house to assess subjective cognitive difficulties related to meal preparation, shopping, remembering verbal information, and planning daily tasks (Jespersen et al., 2024a).

2.4. Statistical analyses

Scores on CAVIR and neuropsychological tests were *z*-transformed based on the mean and SD of HCs. For CAVIR, five cognitive domains were calculated by averaging the *z*-transformed scores within each subtask (Table 1). A cognitive composite score (CAVIR global) was then calculated by averaging the five domains. The same method was used to estimate five cognitive domains and a composite score based on the traditional neuropsychological tests (see Appendix A). Continuous data were assessed for normality using Shapiro-Wilk tests, and non-parametric tests were conducted whenever the assumption of normality was violated. Patients and HCs were compared on demographic and clinical variables using independent sample *t*-tests or Mann-Whitney *U* tests for normally and non-normally distributed data.

Groups were also compared on premorbid intellectual ability estimated from the error score on the DART (Nelson and Willison, 1991). A $\chi 2$ test was applied to investigate group differences in sex distribution (male/female as assigned at birth).

The convergent validity of CAVIR against standard measures of cognition and functioning was assessed with correlation analyses using Pearson's r or Spearman's ρ (r_s). To interpret the strength of the associations, we used the benchmark scale for Pearson's and Spearman's correlation coefficients reported by Dancey and Reidy (2007) (see Table A1 in Appendix A). An r-value $\geq$0.7 was set as the criterion for convergent validity, indicating the instruments measure the same construct (De Vet et al., 2011). Associations were investigated between the global CAVIR and neuropsychological scores and explored between CAVIR subtasks, corresponding neuropsychological domains, and subjective cognitive complaints. We also investigated associations between global CAVIR performance and the AMPS ADL process ability measure, FAST, and UPSA-B. Associations between CAVIR subtasks and ADL process ability were also explored. Significant associations were followed up with post-hoc multiple regression analyses with CAVIR (global or subtask scores), sex, and age as independent variables to investigate if CAVIR performance predicted neuropsychological performance and functioning across sexes and age groups.

The sensitivity of CAVIR was investigated for the global score and explored for each subtask using ANCOVAs with CAVIR scores as the dependent variable, group (patient vs. HC) as fixed factor, and any demographic or clinical variables on which the two groups differed as covariates. Similar post-hoc analyses were conducted to investigate differences in CAVIR global performance separately for MD vs. HC and PD vs. HC and between the participants who received CAVIR version 1 or 2. Finally, using the AMPS ADL employability cut-offs of a motor score < 2.5 logits and a process score < 1.2 logits (Vandamme, 2010), patients were classified as either *not ready for regular employment* (below cut-offs) or *likely ready for regular employment* (above cut-offs). We then compared CAVIR global performance between these two sub-groups.

All analyses were performed using SPSS 25 for Windows. The α-level was set to $p \leq 0.05$ for the primary hypotheses testing and $p \leq 0.01$ for the exploratory analyses investigating the five CAVIR subtasks to control for multiple comparisons (Bonferroni).

3. Results

3.1. Demographics and clinical variables

Demographic and clinical data are presented in Table 2. We included 70 patients with MD or PD (MD = 52, PD = 18) and 70 HCs. Comparisons between patients (MD + PD) and HCs revealed that they were comparable for age, sex, premorbid verbal IQ, and YMRS scores ($p_s \geq$ 0.09). As expected, patients had, despite similar estimated IQ, fewer years of education ($t(138) = 3.06$, $p < 0.01$) and displayed more subsyndromal symptoms of depression, as indicated by higher HDRS scores ($U = 719.0$, $p < 0.001$). Forty-seven patients (MD = 29, PD = 18), all from the VR cognitive remediation trial, were assessed with the AMPS. These patients were comparable to the 23 patients from the hypoxia cognition training trial on all demographic, clinical or cognitive variables ($p_s \geq 0.16$).

Table 2
Demographic and clinical variables, composite scores for Cognition Assessment in Virtual Reality (CAVIR) and neuropsychological tests, functional capacity, and subjective cognition for the patients with mood disorders ($n = 52$) or psychosis spectrum disorders ($n = 18$) and healthy controls ($n = 70$).

Variable	Group		*P*-values (unadjusted)
	MD/PD (N = 70) Mean (SD)	HCs (N = 70) Mean (SD)	
Demographic and clinical variables			
Age	33.4 (10.9)	30.3 (8.9)	0.19
Years of education	14.4 (2.5)	15.6 (2.2)	**< 0.01****
Sex, no. women (%)	50 (71)	43 (61)	0.21
Est. premorbid verbal IQ	111.8 (5.7)	111.4 (5.2)	0.637
HDRS-17 baseline	4.4 (3.8)	0.8 (1.1)	**< 0.001*****
YMRS baseline	1.0 (1.8)	0.4 (1.0)	0.09
BNSS[a]	18.7 (13.1)	–	
SAPS psychotic[b]	1.8 (1.9)	–	
SAPS disorganised	0.5 (1.0)	–	
CAVIR subtasks			
Task 1: Verbal learning/memory	−0.7 (1.3)	0.0 (1.0)	**< 0.001*****
Task 2: Executive functions	−0.2 (1.0)	0.0 (1.0)	0.95
Task 3: Processing speed	−0.8 (1.2)	0.0 (1.0)	**< 0.001*****
Task 4: Working memory	−0.7 (1.4)	0.0 (1.0)	**< 0.001*****
Task 5: Sustaining attention	−1.3 (1.5)	0.0 (1.0)	**< 0.001*****
Global composite score	−0.7 (0.8)	0.0 (0.7)	**< 0.001*****
Neuropsychological domains			
Verbal learning/memory	−0.7 (1.1)	0.0 (0.8)	**< 0.001*****
Executive functions	−0.5 (0.8)	0.0 (0.6)	**< 0.001*****
Processing speed	−0.8 (1.0)	0.0 (0.8)	**< 0.001*****
Working memory	−0.7 (0.8)	0.0 (0.8)	**< 0.001*****
Attention	−0.6 (0.9)	0.0 (0.8)	**< 0.001*****
Global composite score	−0.7 (0.7)	0.0 (0.5)	**< 0.001*****
Measures of functioning			
AMPS motor ADL ability[c]	2.2 (0.3)	–	–
AMPS process ADL ability	1.3 (0.2)	–	–
FAST total	23.5 (11.3)	1.5 (2.5)	**< 0.001*****
UPSA-B total[d]	85.5 (7.0)	89.2 (8.2)	**0.02***
Questionnaires			
Subjective cognition, COBRA	26.9 (7.0)	9.1 (5.3)	**< 0.001*****
Subjective cognition, CODEL[e]	26.6 (7.3)	8.8 (4.7)	**< 0.001*****

Abbreviations and notes. MD: mood disorder; PD: psychosis spectrum disorder; HC: Healthy control. HDRS-17: 17 item Hamilton Depression Rating Scale. YMRS: Young Mania Rating Scale; BNSS: Brief Negative Symptom Scale. SAPS: Scale for the Assessment of Positive Symptoms; CAVIR: Cognition Assessment in VItual reality; AMPS: the Asssessment of Motor and Process Skills; ADL: Activities of Daily Living; FAST: Functional Assessment Short Test. UPSA-B: University of California, San Diego (UCSD) Performance-based Skills Assessment - Brief. COBRA: Cognitive Complaints in Bipolar Disorder Rating Assessment; CODEL: Cognitive Difficulties in Everyday Life; M:Mean; SD: Standard Deviation. * = $p < 0.05$ (two-tailed), ** = $p \leq 0.01$ (two-tailed), *** = $p < 0.001$ (two-tailed). [a]Data for BNSS was only collected for patients with PD (n = 18); [b]Data for SAPS was only collected for patients with PD (n = 18); [c]Data for the AMPS was only collected for patients in the virtual reality cognitive remediation trial (MD = 29; PD = 18); [d]Data for the UPSA-B was only collected for participants in the virtual reality cognitive remediation trial (HC = 33; MD = 29; PD = 18); [e]Data for the CODEL was only collected for participants in the virtual reality cognitive remediation trial (HC = 33; MD = 29; PD = 18).

3.2. Association between CAVIR and standard measures of cognition

Analyses across the entire sample showed a moderate positive correlation between the global CAVIR and global neuropsychological test scores ($r_s(138) = 0.60$, $p < 0.001$; Fig. 2A and Table 3). Adjusting for age and sex revealed a significant model ($R^2 = 0.43$, $F(3,136) = 33.86$, $p <$ 0.001), with the global CAVIR score being the only significant predictor of neuropsychological performance ($\beta = 0.67$, $p < 0.001$; for remaining predictors $p_s \geq 0.50$). Performance on CAVIR subtask 3 correlated moderately with the corresponding neuropsychological domain of processing speed ($r(138) = 0.57$, $p < 0.001$; Table 3). For the remaining CAVIR subtasks, analyses revealed significant but weak associations with the corresponding neuropsychological domain for verbal memory ($r_s(138) = 0.23$, $p < 0.01$), executive function ($r(138) = 0.22$, $p < 0.01$), working memory ($r_s(138) = 0.25$, $p < 0.01$), and attention ($r_s(138) =$ 0.29, $p < 0.001$; Table 3). These associations prevailed after adjusting for sex and age ($p_s \leq 0.01$).

Across the whole sample, lower CAVIR global scores showed weak associations with more subjective cognitive complaints on the COBRA ($r_s(138) = -0.35$, $p < 0.001$) and CODEL ($r_s(78) = -0.32$, $p < 0.01$),

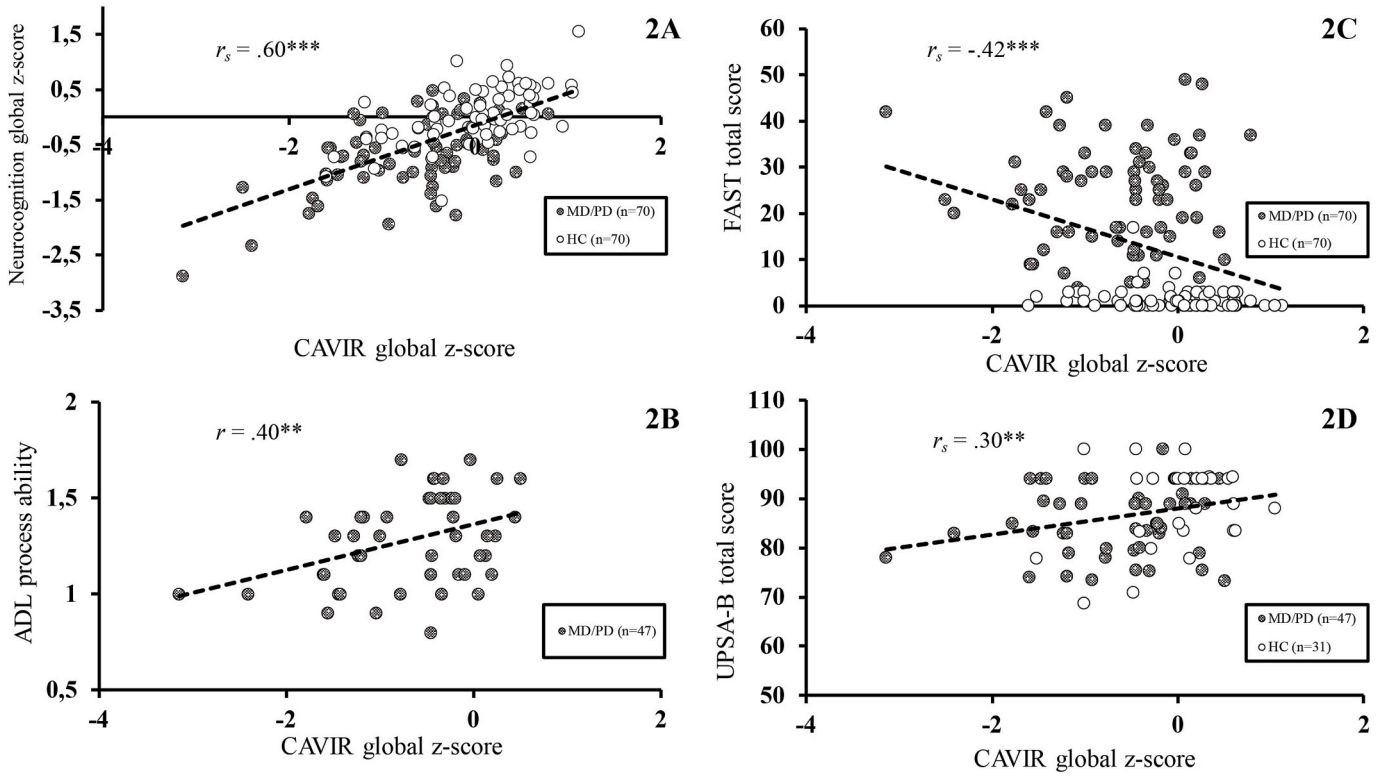

Fig. 2. Left panel: Correlations between global CAVIR and neuropsychological performance (2 A) and ADL process ability (2B). Right panel: Correlations between CAVIR global score and FAST (2C) and UPSA-B (2D). Abbreviations: MD: mood disorder; PD: psychosis spectrum disorder; HC: healthy control; CAVIR: Cognition Assessment in Virtual Reality; ADL: Activities of Daily Living, FAST: Functional Assessment Short Test; UPSA-B: University of California, San Diego Performance-based Skills Assessment - Brief. ** = $p \leq 0.01$ (two-tailed), *** = $p < 0.001$ (two-tailed).

Table 3
Correlations between CAVIR and neuropsychological/functional measures.

	CAVIR global composite score and subtasks					
	CAVIR task 1: verbal memory	CAVIR task 2: executive function	CAVIR task 3: processing speed	CAVIR task 4: working memory	CAVIR task 5: sustained attention	CAVIR global composite score
	r/r_s	r/r_s	r/r_s	r/r_s	r/r_s	r/r_s
Neuropsychological domains						
Verbal learning and memory (n = 140)	**0.23****	0.14	**0.33*****	**0.31*****	**0.27*****	**0.39*****
Executive functions (n = 140)	**0.33*****	**0.22****	**0.51*****	**0.36*****	**0.44*****	**0.57*****
Processing speed (n = 140)	**0.32*****	0.13	**0.57*****	**0.27****	**0.32*****	**0.46*****
Working memory (n = 140)	**0.23****	0.16	**0.40*****	**0.25****	**0.50*****	**0.46*****
Attention (n = 140)	**0.23****	0.14	**0.34*****	**0.30*****	**0.29*****	**0.37*****
Neuropsychological global composite score (n = 140)	**0.36*****	**0.21***	**0.56*****	**0.39*****	**0.48*****	**0.60*****
Functional capacity measures						
FAST (n = 140)	**−0.27*****	−0.03	**−0.28*****	**−0.37*****	**−0.41*****	**−0.41*****
UPSA-B (n = 80)	**0.25***	0.10	0.19	0.10	**0.32****	**0.30****
AMPS Activities of Daily Living (ADL) ability						
ADL motor ability (n = 47)	0.01	0.15	0.15	0.26	0.09	0.18
ADL process ability (n = 47)	0.12	**0.38****	0.20	**0.38****	0.15	**0.40****
Subjective cognitive complaints						
COBRA (n = 140)	**−0.24****	−0.05	**−0.24****	**−0.34****	**−0.36*****	**−0.35*****
CODEL (n = 80)	−0.17	−0.03	−0.20	**−0.25****	**−0.32*****	**−0.32****

Abbreviations and notes: CAVIR: Cognition Assessement in Vitual reality; AMPS: the Assessment of Motor and Process Skills; ADL: Activities of Daily Living; FAST: Functional Assessment Short Test. UPSA-B: University of California, San Diego (UCSD) Performance-based Skills Assessment - Brief. COBRA: Cognitive Complaints in Bipolar Disorder Rating Assessment; CODEL: Cognitive Difficulties in Everyday Life. * = $p < 0.05$ (two-tailed), ** = $p \leq 0.01$ (two-tailed), *** = $p < 0.001$ (two-tailed). For the five CAVIR subtasks, only the associations marked with ** or *** were statistically significant after Bonferroni correction (0.05/5 = 0.01).

which survived adjustment for age and sex ($p_s \leq 0.02$). A sub-analysis for those with available data on both the COBRA and CODEL questionnaires revealed a strong association between self-ratings of cognitive difficulties on these instruments ($r_s(78) = 0.92$, $p < 0.001$).

3.3. Association between CAVIR and measures of functioning

There was a significant weak to moderate association between higher CAVIR global scores and better AMPS ADL *process* ability in the patient subsample ($r(45) = 0.40$, $p < 0.01$; Fig. 2B and Table 3). Adjusting for

age and sex revealed a significant model ($R^2 = 0.18$, $F(3,43) = 3.01$, $p = 0.04$), with global CAVIR performance being the only significant predictor of ADL process ability ($\beta = 0.37$, $p = 0.03$; for remaining predictors $p_s \geq 0.27$). Analyses showed significant weak associations between ADL process ability and CAVIR subtasks 2 ($r(45) = 0.38$, $p < 0.01$) and 4 ($r_s(45) = 0.38$, $p < 0.01$; for remaining subtasks $p_s \geq 0.11$; Table 3). The association between CAVIR subtask 2 and ADL process ability survived adjustment for sex and age ($\beta = 0.37$, $p < 0.01$), while the association with subtask 4 was reduced to a trend ($p = 0.09$).

In comparison, no significant associations were observed between patients' ADL process ability and neuropsychological performance (neither global nor subdomain scores; $p_s \geq 0.09$; Table B1 in Appendix B). Similarly, there were no significant associations between patients' ADL process ability and total scores on the FAST, UPSA-B, COBRA, and CODEL ($p_s \geq 0.30$; Table B1).

Across the entire sample, *lower* global CAVIR performance was significantly associated with more functional disability as reflected by a moderate correlation with *higher* scores on the FAST ($r_s(138) = -0.42$, $p < 0.001$; Fig. 2C) and a weak correlation with *lower* scores on the UPSA-B ($r_s(76) = 0.30$, $p < 0.01$; Fig. 2D), which prevailed after adjusting for age and sex ($p_s \leq 0.01$). Lower global neuropsychological performance was also weakly to moderately associated with higher FAST scores and lower UPSA-B scores ($p_s \leq 0.01$; Table B1 in Appendix B).

3.4. Sensitivity of CAVIR

The patient group showed impaired performance on the CAVIR global score compared to HCs with a large effect size also after controlling for years of education and HDRS scores ($F(1,138) = 27.92$, $p < 0.001$, $\eta_p^2 = 0.17$; Table 2 and Fig. 3). For CAVIR subtasks, controlling for education and HDRS scores, the patient group displayed impaired performance relative to the HC group on subtask 1 ($F(1,138) = 12.06$, $p < 0.001$, $\eta_p^2 = 0.08$), subtask 3 ($F(1,138) = 14.83$, $p < 0.001$, $\eta_p^2 = 0.10$), subtask 4 ($F(1,138) = 11.11$, $p < 0.001$, $\eta_p^2 = 0.07$), and subtask 5 ($F(1,138) = 25.27$, $p < 0.001$, $\eta_p^2 = 0.16$). In contrast, there was no difference between the groups on CAVIR subtask 2 ($p = 0.95$; Table 2 and Fig. 3).

A post-hoc analysis showed impaired CAVIR global performance in patients with MD compared to HCs, also after controlling for years of education, HDRS-, and YMRS scores ($F(1,120) = 23.15$, $p < 0.001$, $\eta_p^2 = 0.17$; Fig. B1 in Appendix B). Patients with PD also showed impaired performance on the CAVIR global score compared to HCs, which prevailed after controlling for age, education, and HDRS scores ($F(1,86) = 8.91$, $p < 0.01$, $\eta_p^2 = 0.10$; Fig. B1).

Post-hoc analyses across the entire sample revealed no differences in global CAVIR performance between the participants who completed CAVIR version 1 and 2, respectively ($p = 0.43$). In concordance, these two groups showed no differences in global neuropsychological performance ($p = 0.24$).

3.5. Occupational capacity and CAVIR performance

Of the 47 patients who underwent the AMPS, 34 % ($n = 16$) obtained an ADL motor- and process ability score *below* the cut-offs of <2.5 and < 1.2 logits, suggesting they were not ready for regular employment. In contrast, 66 % ($n = 31$) scored *above* these cut-offs, suggesting they were likely ready for regular employment. The division of patients into these two ADL groups was mirrored by a difference in their real-life employment status: Among those who scored *below* and *above* the cut-offs, 87 % ($n = 14$) and 55 % ($n = 17$) were unemployed or worked/studied part-time, respectively ($X^2(1, 47) = 5.01$, $p = 0.03$). Further, CAVIR test performance differed between these groups with a large effect size; patients who scored *below* the ADL cut-offs displayed clinically significant lower global CAVIR scores (mean *z*-score = −1.1) than those above the ADL cut-offs (mean *z*-score = −0.5; $t(45) = 2.12$, $p = 0.05$, $d = 0.75$). In comparison, there were no significant differences between these two sub-groups on clinical or demographic variables, FAST, UPSA-B, subjective cognition, or global neuropsychological performance ($p_s \geq 0.07$).

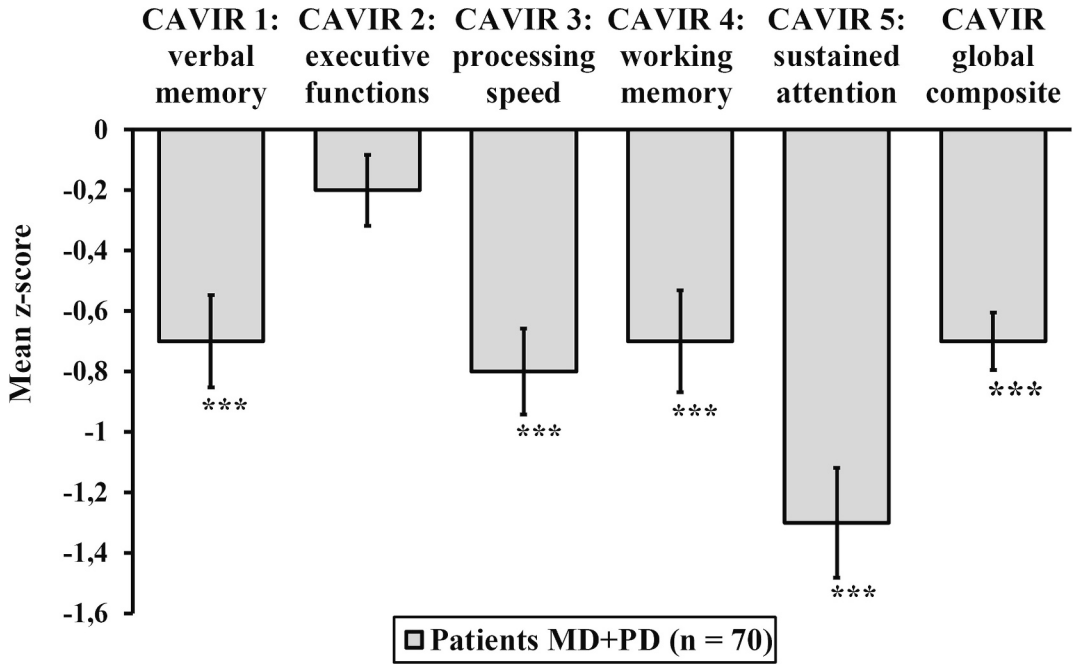

Fig. 3. Difference in performance (before any adjustments) on the five CAVIR subtasks and global composite score between the HC group and the patient (MD + PD) group. The Y-axis denotes the mean unadjusted cognition z-score for the patient group based on the mean and standard deviation (SD) of the HC group. The HC group has a mean of 0, and the bars represent the patient group's deviation from the HC group. Error bars represent SEM. *** = p < 0.001 (two-tailed). For the five CAVIR subtasks, the group differences marked with *** were statistically significant after Bonferroni correction (0.05/5 = 0.01). Abbreviations: MD: mood disorder; PD: psychosis spectrum disorder; HC: healthy control; CAVIR: Cognition Assessment in Virtual Reality; ADL: Activities of Daily Living, FAST: Functional Assessment Short Test; UPSA-B: University of California, San Diego Performance-based Skills Assessment - Brief.

4. Discussion

This study investigated the validity and sensitivity of the optimised Cognition Assessment in Virtual Reality (CAVIR) test in 70 patients with mood or psychosis spectrum disorders and 70 healthy control participants. Our findings provide novel insight into how this immersive head-mounted (HMD) VR cognition test relates to neuropsychological performance and real-life ADL ability in these populations. Consistent with the first hypothesis, better global CAVIR performance was moderately correlated with higher global neuropsychological test scores. In accordance with the second hypothesis, analyses revealed a significant weak to moderate positive association between global CAVIR performance and AMPS evaluation of patients' ADL process ability related to household tasks. Notably, the observed associations between CAVIR and neuropsychological/ADL ability were not influenced by age and sex. In line with the third hypothesis, CAVIR showed high sensitivity for cognitive impairments in MD and PD and could effectively determine patients' occupational capacity.

The moderate positive association between global performance on CAVIR and standard neuropsychological tests is consistent with our first validation study of the prior CAVIR version (Miskowiak et al., 2022b) and previous studies investigating the use of immersive HMD VR tests for cognitive functions in patients with MD (Voinescu et al., 2023) or PD (Adriasola et al., 2024; Mannan et al., 2023). Compared with previous studies, it is a strength that we also investigated the ecological validity of CAVIR by assessing the association with performance-based ADL ability. While most studies on immersive HMD VR cognition tools consider *verisimilitude*, i.e., designing simulations representative of daily-life situations, they often neglect the *veridicality* of the results, i.e., the investigation of how well the tasks predict real-world performance in these situations (Franzen and Wilhelm, 1996; Kirkham et al., 2024). In this context, it is of interest that CAVIR showed a significant association with neuropsychological performance and ADL process ability, as this indicates that CAVIR is related to both underlying cognitive functions and real-world functional skills. Nevertheless, the observed associations were moderate at best, and none met the stringent criteria for convergent validity ($r \geq 0.7$). Therefore, our results suggest that CAVIR is not a test of neuropsychological performance or ADL ability per se but may instead measure a unique, albeit closely related construct, namely *cognitive skills*.

From a theoretical perspective, CAVIR may bridge the assessment of cognitive- and real-world functioning because it combines *construct-driven* and *function-led* test components (Burgess et al., 2006; Parsons et al., 2017). While construct-driven VR tests elucidate performance within specific cognitive domains but lack predictive accuracy regarding real-world functioning, function-led tests evaluate how effectively individuals manage actual everyday challenges, irrespective of the cognitive domains involved (Burgess et al., 2006; Parsons et al., 2017). In keeping with this distinction, it was interesting that particularly CAVIR subtasks 2 and 4 (planning meal preparation and remembering locations of kitchen utensils, respectively) were associated with ADL ability, as both subtasks are primarily *function-led* and target actual real-life skills that are also assessed in the AMPS. In contrast, CAVIR subtasks 1, 3, and 5 showed no significant association with ADL ability, which could be explained by these subtasks being more construct-driven, resembling classic neuropsychological test paradigms embedded in a VR environment. In line with this, the most construct-driven CAVIR task, subtask 3, exhibited the strongest association with neuropsychological performance. Nevertheless, these subtasks still contain function-led elements, such as the need to remember, interact with, and select kitchen items and to maintain focus on the cooking process in the presence of various distracting naturalistic stimuli (e.g., phone notifications, people talking). This integration of real-life elements increases the ecological validity compared to standard neuropsychological tests. Notably, we previously found that CAVIR subtask 5 (sustaining attention while cooking) was highly sensitive to capturing daily-life attentional difficulties experienced by patients with BD and comorbid ADHD (Jespersen et al., 2024b). This highlights the potential utility of this subtask to identify and address specific cognitive deficits related to attentional control in everyday scenarios.

Taken together, we propose that CAVIR elucidate the extent to which a person can translate their cognitive functioning into actual skills in the typical daily activity of meal planning and preparation. In keeping with WHO's ICF framework, CAVIR may provide much-needed insight into patients' *cognitive skills* necessary for tackling daily life challenges (i.e., the ICF level of activity/participation) (WHO, 2013). If the current findings are replicated in further validation studies, CAVIR could become an attractive addition to a traditional neuropsychological assessment battery in studies investigating pro-cognitive treatments, as it may aid insight into whether the effect of interventions translates into real-world functional improvements, which is an important requirement for their approval (Miskowiak et al., 2017). In this context, it is promising that the absence of difference between participants' performance on CAVIR versions 1 and 2 mirrored the absence of difference in neuropsychological performance, as this indicates that the two CAVIR versions are of equal difficulty and can be used for repeated testing. In a clinical setting, CAVIR could offer a feasible and easily implementable tool for healthcare professionals to assess patients' cognitive skills in real-world situations. This assessment could serve as a valuable basis for assisting patients in compensating for or training their cognitive skills in daily life situations, aligning with recommendations from the ISBD Targeting Cognition Task Force (Miskowiak et al., 2017). In this context, it was an important finding that the patients classified as not ready for regular employment based on the ADL ability cut-off also showed an average global performance on CAVIR of 1.1 SD or more below the mean of HCs. This indicates that a score of >1 SD below that of healthy age- and IQ-matched controls may be a relevant cut-off for estimating patients' occupational capacity, subject to replication in a larger-scale study. Another interesting observation was that CAVIR subtask 2 (planning meal preparation) showed the closest association to ADL process ability. However, subtask 2 was the only subtask on which patients and HCs showed no performance difference, indicating that it is highly challenging for both groups. Given its significant, albeit weak, association with ADL process ability, this subtask may also be relevant for measuring daily-life cognitive skills in non-clinical samples or individuals with non-psychiatric conditions who experience cognitive decline.

In comparison, neither global neuropsychological performance, subjective cognition, FAST, nor UPSA-B scores correlated with ADL process ability, although we did observe a trend towards a significant weak association for the executive function domain. Our results corroborate previous findings in MD and PD of weak associations between AMPS ADL ability and neuropsychological performance, particularly for processing speed and executive functions (Gildengers et al., 2007; Josman and Katz, 2006; Träger et al., 2017). As for self-report measures of cognition and functioning (i.e., questionnaires or interviews), studies have generally found no or weak associations with ADL ability in psychiatric populations (Nielsen and Wæhrens, 2015; Träger et al., 2017) and other diagnostic groups (Amris et al., 2022; Wæhrens et al., 2012). Our results are in accordance with these findings, suggesting that performance-based and self-report measures assess different aspects of cognition and functioning. As previously suggested, this discrepancy may partly be explained by patients' experience of cognitive impairment being influenced by subclinical depressive symptoms, introspective capacity, and expectations in the assessment situation (Nielsen and Wæhrens, 2015; Petersen et al., 2019; Träger et al., 2017). Together, this highlights the importance of objective tests that provide insight into daily-life cognitive and functional abilities without being influenced by psychiatric symptoms and introspective abilities. Applying performance-based ADL measures such as AMPS or other real-world measures, such as the Multiple Errands Test (MET), could aid the achievement of new insight into the transfer effects of cognitive

interventions (Torralva et al., 2012; Träger et al., 2017). Nevertheless, these tests require a specific real-world environment, such as a standardised test apartment, home visit, or shopping mall. They also necessitate the presence of a certified occupational therapist or examiner to accompany the participant, which limits their feasibility as outcome measures in pro-cognitive treatment trials. In comparison, CAVIR is quite cost-effective as it is largely self-administered and involves affordable VR equipment. From a practical perspective, CAVIR is thus an easily implementable performance-based measure of daily life cognitive skills in both clinical settings and treatment trials pending further validation.

A strength of the study was that patients were symptomatically stable, which enables insight into more persistent impairments that are not directly caused by acute mood- or psychotic symptoms. While the relatively large sample size ($n = 140$) is a strength, only a few patients had a psychosis spectrum diagnosis ($n = 18$), limiting the power for determining the validity of CAVIR for separate diagnostic groups. Therefore, our findings warrant further validation studies of CAVIR in larger diagnosis-specific samples or transdiagnostic samples stratified by diagnosis. Also, only a modest sub-sample of patients ($n = 47$) completed the AMPS, and it could be considered a limitation that patients were not assessed in their actual homes, although items in the test apartment were arranged to mimic their home environment. However, this could also be considered a strength as it allowed us to keep a controlled test environment. Another limitation was that we did not investigate the potential influence of medication on CAVIR performance. Finally, it should be noted that around half of the HC participants were blood donors who have been shown to be healthier than the general background population (Brodersen et al., 2023).

In conclusion, the immersive VR cognition assessment tool, CAVIR, is a sensitive measure of cognitive impairment in mood and psychosis spectrum disorders that is associated with neuropsychological performance and ADL ability related to household tasks. This highlights CAVIR as a promising instrument for bridging standard cognitive assessment and real-world functioning to better target patients' daily-life cognitive skills. Work is now underway to further assess the cut-off on the CAVIR with optimal sensitivity and specificity for cognitive impairment and its test-retest reliability and to validate translated versions in English, German, and Spanish in collaboration with members of the ISBD Targeting Cognition Task Force. Future studies are warranted to replicate the ability of CAVIR to determine patients' vocational abilities and to examine the validity and feasibility of implementing CAVIR in clinical settings and pro-cognitive treatment trials.

Ethics approval and consent to participate

The virtual reality-based cognitive remediation trial and hypoxia cognition training trial has both been approved by the Committee on Health Research Ethics in the Capital Region of Denmark (protocol number: H-22004153 and H-22028111, respectively) and by the Danish Data Protection Agency at the Capital Region of Denmark (protocol number: P-2022-411 and P-2022-354, respectively). Both studies are conducted in compliance with existing laws on data protection. Participants are given written and verbal information about the studies to make an informed decision about their participation. Written informed consent is obtained from all participants using the approved consent form template from the Committee on Health Research Ethics in the Capital Region of Denmark. The study was conducted in accordance with the recommendations of the Declaration of Helsinki.

Funding

The data collection for the VR cognitive remediation Trial was supported by TrygFonden (grant no. 150128), the Axel Muusfeldts Foundation (grant no. 2022-0097), and the Jascha Foundation (grant no. 2022-0134). The data collection for the Altitude-like training trial was funded by a European Research Council (ERC) Consolidator Grant (grant no. 101043416). The funding bodies had no role in the data collection, analysis, or interpretation of data, in the writing of the article, or in the decision to submit the article for publication.

CRediT authorship contribution statement

Andreas E. Jespersen: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Investigation, Formal analysis, Data curation. **Anders Lumbye:** Writing – review & editing, Software. **Johanna Schandorff:** Writing – review & editing, Project administration, Investigation, Data curation. **Viktoria Damgaard:** Writing – review & editing, Project administration, Investigation, Data curation. **Louise B. Glenthøj:** Writing – review & editing, Resources. **Merete Nordentoft:** Writing – review & editing, Resources. **Christina Mikkelsen:** Writing – review & editing, Resources. **Maria Didriksen:** Writing – review & editing, Resources. **Sisse R. Ostrowski:** Writing – review & editing, Resources. **Maj Vinberg:** Writing – review & editing, Resources, Methodology. **Eva E. Wæhrens:** Writing – review & editing, Resources, Methodology, Investigation, Data curation. **Kamilla W. Miskowiak:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

Kamilla Miskowiak has received consultancy fees from Lundbeck, Angelini, Gedeon Richter, and Janssen-Cilag in the past four years, and Maj Vinberg has received consultancy fees from Lundbeck and Janssen-Cilag within the past three years. The remaining authors report no conflicts of interest.

Acknowledgements

The authors wish to thank the Copenhagen Affective Disorder Research Clinic and the early intervention clinics for psychotic disorders (OPUS), as well as other mental health centres in the Capital Region of Denmark, for their assistance with recruiting participants. Occupational therapists Maria Cecilie Von Bülow, Mie Skovmand Christensen, Masumi Takeuchi Holm, Thomas Buhl-Wiggers, and Signe Johanne Berg Bojesen at the Parker Institute, Frederiksberg Hospital are acknowledged for contributing to the completion and scoring of the AMPS assessments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2024.10.095.

References

Adriasola, A., Torres, S.C., Cañada, Y., Chicchi Giglioli, I.A., García-Blanco, A., Sierra, P., López-Cerveró, M., Chloe, B.R., Navalón, P., Mariano, A.R., 2024. Assessing executive functioning in schizophrenia: concurrent and discriminative validity of a novel virtual cooking task. Cyberpsychol. Behav. Soc. Netw. https://doi.org/10.1089/cyber.2023.0443.

Amris, K., Bandak, E., Kristensen, L.E., Wæhrens, E.E., 2022. Agreement between self-reported and observed functioning in patients with rheumatoid arthritis, osteoarthritis, and fibromyalgia, and the influence of pain and fatigue: a cross-sectional study. Scand. J. Rheumatol. 51 (6), 452–460. https://doi.org/10.1080/03009742.2021.1952755.

Andreasen, N.C., 1984. Scale for the Assesement of Negative Positive Symptoms (SAPS). University of Iowa, Iowa City.

Army Individual Test Battery, 1944. Manual of Directions and Scoring. War Department, Adjutant General's Office, Washington, DC.

Ayres, H., John, A.P., 2015. The assessment of motor and process skills as a measure of ADL ability in schizophrenia. Scand. J. Occup. Ther. 22 (6), 470–477. https://doi.org/10.3109/11038128.2015.1061050.

Bohil, C.J., Alicea, B., Biocca, F.A., 2011. Virtual reality in neuroscience research and therapy. Nat. Rev. Neurosci. 12 (12), 752–762. https://doi.org/10.1038/nrn3122.
Bora, E., Pantelis, C., 2015. Meta-analysis of cognitive impairment in first-episode bipolar disorder: comparison with first-episode schizophrenia and healthy controls. Schizophr. Bull. 41 (5), 1095–1104. https://doi.org/10.1093/schbul/sbu198.
Borkowski, J., Benton, A., Spreen, O., 1967. Word fluency and brain damage. Neuropsychologia 5, 5.
Bourne, C., Aydemir, O., Balanza-Martinez, V., Bora, E., Brissos, S., Cavanagh, J.T., Clark, L., Cubukcuoglu, Z., Dias, V.V., Dittmann, S., Ferrier, I.N., Fleck, D.E., Frangou, S., Gallagher, P., Jones, L., Kieseppa, T., Martinez-Aran, A., Melle, I., Moore, P.B., et al., 2013. Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis. Acta Psychiatr. Scand. 128 (3), 149–162. https://doi.org/10.1111/acps.12133.
Bowie, C.R., Reichenberg, A., Patterson, T.L., Heaton, R.K., Harvey, P.D., 2006. Determinants of real-world functional performance in schizophrenia subjects: correlations with cognition, functional capacity, and symptoms. Am. J. Psychiatry 163 (3), 418–425. https://doi.org/10.1176/appi.ajp.163.3.418.
Brodersen, T., Rostgaard, K., Lau, C.J., Juel, K., Erikstrup, C., Nielsen, K.R., Ostrowski, S. R., Titlestad, K., Saekmose, S.G., Pedersen, O.B.V., Hjalgrim, H., 2023. The healthy donor effect and survey participation, becoming a donor and donor career. Transfusion 63 (1), 143–155. https://doi.org/10.1111/trf.17190.
Burgess, P.W., Alderman, N., Forbes, C., Costello, A., Coates, L.M., Dawson, D.R., Anderson, N.D., Gilbert, S.J., Dumontheil, I., Channon, S., 2006. The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. J. Int. Neuropsychol. Soc. 12 (2), 194–209. https://doi.org/10.1017/s1355617706060310.
Comment. WHOEdf, October 2013. How to use the ICF: A practical manual for using the International Classification of Functioning, Disability and Health (ICF). WHO, Geneva.
Crawford, J.R., Besson, J.A., Parker, D.M., Sutherland, K.M., Keen, P.L., 1987. Estimation of premorbid intellectual status in depression. Br. J. Clin. Psychol. 26 (4), 313–314. https://doi.org/10.1111/j.2044-8260.1987.tb01366.x.
Dancey, C.P., Reidy, J., 2007. Statistics without Maths for Psychology. Pearson Education.
De Vet, H.C.W., Terwee, C.B., Mokkink, L.B., Knol, D.L., 2011. Measurement in Medicine: A Practical Guide. Cambridge University Press. https://doi.org/10.1017/CBO9780511996214.
Fisher, Anne G., K. B. J., 2012. Assessment of Motor and Process Skills. Vol. 1: development, Standardization, and Administration Manual.
Franzen, M.D., Wilhelm, K.L., 1996. Conceptual foundations of ecological validity in neuropsychological assessment. In: Ecological Validity of Neuropsychological Testing. Gr Press/St Lucie Press, Inc., pp. 91–112
Gildengers, A.G., Butters, M.A., Chisholm, D., Rogers, J.C., Holm, M.B., Bhalla, R.K., Seligman, K., Dew, M.A., Reynolds 3rd, C.F., Kupfer, D.J., Mulsant, B.H., 2007. Cognitive functioning and instrumental activities of daily living in late-life bipolar disorder. Am. J. Geriatr. Psychiatry 15 (2), 174–179. https://doi.org/10.1097/JGP.0b013e31802dd367.
Goldberg, Z., Kuslak, B., Kurtz, M.M., 2023. A meta-analytic investigation of cognitive remediation for mood disorders: efficacy and the role of study quality, sample and treatment factors. J. Affect. Disord. 330, 74–82. https://doi.org/10.1016/j.jad.2023.02.137.
Green, M.F., Kern, R.S., Braff, D.L., Mintz, J., 2000. Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"? Schizophr. Bull. 26 (1), 119–136. https://doi.org/10.1093/oxfordjournals.schbul.a033430.
Green, M.J., Girshkin, L., Kremerskothen, K., Watkeys, O., Quidé, Y., 2020. A systematic review of studies reporting data-driven cognitive subtypes across the psychosis spectrum. Neuropsychol. Rev. 30 (4), 446–460. https://doi.org/10.1007/s11065-019-09422-7.
Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 23, 56–62. https://doi.org/10.1136/jnnp.23.1.56.
Haslam, J., Pépin, G., Bourbonnais, R., Grignon, S., 2010. Processes of task performance as measured by the Assessment of Motor and Process Skills (AMPS): a predictor of work-related outcomes for adults with schizophrenia? Work 37 (1), 53–64. https://doi.org/10.3233/wor-2010-1056.
Hørlyck, L.D., Obenhausen, K., Jansari, A., Ullum, H., Miskowiak, K.W., 2021. Virtual reality assessment of daily life executive functions in mood disorders: associations with neuropsychological and functional measures. J. Affect. Disord. 280 (Pt A), 478–487. https://doi.org/10.1016/j.jad.2020.11.084.
Jaeger, J., Vieta, E., 2007. Functional outcome and disability in bipolar disorders: ongoing research and future directions. Bipolar Disord. 9 (1–2), 1–2. https://doi.org/10.1111/j.1399-5618.2007.00441.x.
Jaeger, J., Berns, S., Uzelac, S., Davis-Conway, S., 2006. Neurocognitive deficits and disability in major depressive disorder. Psychiatry Res. 145 (1), 39–48. https://doi.org/10.1016/j.psychres.2005.11.011.
Jespersen, A.E., Lumbye, A., Vinberg, M., Glenthøj, L., Nordentoft, M., Wæhrens, E.E., Knudsen, G.M., Makransky, G., Miskowiak, K.W., 2024a. Effect of immersive virtual reality-based cognitive remediation in patients with mood or psychosis spectrum disorders: study protocol for a randomized, controlled, double-blinded trial. Trials 25 (1), 82. https://doi.org/10.1186/s13063-024-07910-7.
Jespersen, A.E., Obel, Z., Lumbye, A., Kessing, L.V., Miskowiak, K.W., 2024b. Bipolar-ADHD comorbidity: screening for differences in neurocognition and virtual reality-based cognitive performance. Nord. J. Psychiatry 78 (3), 238–246. https://doi.org/10.1080/08039488.2024.2309496.
Josman, N., Katz, N., 2006. Relationships of categorization on tests and daily tasks in patients with schizophrenia, post-stroke patients and healthy controls. Psychiatry Res. 141 (1), 15–28. https://doi.org/10.1016/j.psychres.2004.03.015.
Keefe, R.S.E., Davis, V.G., Atkins, A.S., Vaughan, A., Patterson, T., Narasimhan, M., Harvey, P.D., 2016. Validation of a computerized test of functional capacity. Schizophr. Res. 175 (1–3), 90–96. https://doi.org/10.1016/j.schres.2016.03.038.
Kirkham, R., Kooijman, L., Albertella, L., Myles, D., Yücel, M., Rotaru, K., 2024. Immersive Virtual Reality–Based Methods for Assessing Executive Functioning: Systematic Review. JMIR Serious Games 12, e50282. https://doi.org/10.2196/50282.
Kirkpatrick, B., Strauss, G.P., Nguyen, L., Fischer, B.A., Daniel, D.G., Cienfuegos, A., Marder, S.R., 2011. The brief negative symptom scale: psychometric properties. Schizophr. Bull. 37 (2), 300–305. https://doi.org/10.1093/schbul/sbq059.
Makransky, G., Petersen, G.B., 2021. The Cognitive Affective Model of Immersive Learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. Educ. Psychol. Rev. 33 (3), 937–958. https://doi.org/10.1007/s10648-020-09586-2.
Makransky, G., Terkildsen, T.S., Mayer, R.E., 2019. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learn. Instr. 60, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007.
Mannan, F.A., Porffy, L.A., Joyce, D.W., Shergill, S.S., Celiktutan, O., 2023. Automatic detection of cognitive impairment with virtual reality. Sensors (Basel) 23 (2). https://doi.org/10.3390/s23021026.
Mausbach, B.T., Harvey, P.D., Goldman, S.R., Jeste, D.V., Patterson, T.L., 2007. Development of a brief scale of everyday functioning in persons with serious mental illness. Schizophr. Bull. 33 (6), 1364–1372. https://doi.org/10.1093/schbul/sbm014.
Mausbach, B.T., Harvey, P.D., Pulver, A.E., Depp, C.A., Wolyniec, P.S., Thornquist, M.H., Luke, J.R., McGrath, J.A., Bowie, C.R., Patterson, T.L., 2010. Relationship of the Brief UCSD Performance-based Skills Assessment (UPSA-B) to multiple indicators of functioning in people with schizophrenia and bipolar disorder. Bipolar Disord. 12 (1), 45–55. https://doi.org/10.1111/j.1399-5618.2009.00787.x.
Miskowiak, K.W., Burdick, K., Martínez-Arán, A., Bonnin, C., Bowie, C., Carvalho, A., Gallagher, P., Lafer, B., Lopez Jaramillo, C., Sumiyoshi, T., McIntyre, R.S., Schaffer, A., Porter, R., Torres, I., Yatham, L., Young, A., Kessing, L., Vieta, E., 2017. Methodological recommendations for cognition trials in bipolar disorder by the International Society for Bipolar Disorders Targeting Cognition Task Force. Bipolar Disord. 19. https://doi.org/10.1111/bdi.12534.
Miskowiak, K.W., Burdick, K.E., Martinez-Aran, A., Bonnin, C.M., Bowie, C.R., Carvalho, A.F., Gallagher, P., Lafer, B., Lopez-Jaramillo, C., Sumiyoshi, T., McIntyre, R.S., Schaffer, A., Porter, R.J., Purdon, S., Torres, I.J., Yatham, L.N., Young, A.H., Kessing, L.V., Vieta, E., 2018. Assessing and addressing cognitive impairment in bipolar disorder: the International Society for Bipolar Disorders Targeting Cognition Task Force recommendations for clinicians. Bipolar Disord. 20 (3), 184–194. https://doi.org/10.1111/bdi.12595.
Miskowiak, K.W., Seeberg, I., Jensen, M.B., Balanzá-Martínez, V., Del Mar Bonnin, C., Bowie, C.R., Carvalho, A.F., Dols, A., Douglas, K., Gallagher, P., Hasler, G., Lafer, B., Lewandowski, K.E., López-Jaramillo, C., Martinez-Aran, A., McIntyre, R.S., Porter, R. J., Purdon, S.E., Schaffer, A., et al., 2022a. Randomised controlled cognition trials in remitted patients with mood disorders published between 2015 and 2021: a systematic review by the International Society for Bipolar Disorders Targeting Cognition Task Force. Bipolar Disord. 24 (4), 354–374. https://doi.org/10.1111/bdi.13193.
Miskowiak, K.W., Jespersen, A.E., Kessing, L.V., Aggestrup, A.S., Glenthøj, L.B., Nordentoft, M., Ott, C.V., Lumbye, A., 2022b. Cognition assessment in virtual reality: validity and feasibility of a novel virtual reality test for real-life cognitive functions in mood disorders and psychosis spectrum disorders. J. Psychiatr. Res. 145, 182–189. https://doi.org/10.1016/j.jpsychires.2021.12.002.
Neguț, A., Matu, S.A., Sava, F.A., David, D., 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.
Nelson, H.E., Willison, J., 1991. National Adult Reading Test (NART). Nfer-Nelson Windsor.
Nielsen, K.T., Wæhrens, E.E., 2015. Occupational therapy evaluation: use of self-report and/or observation? Scand. J. Occup. Ther. 22 (1), 13–23. https://doi.org/10.3109/11038128.2014.961547.
Østergaard Christensen, T., Vesterager, L., Krarup, G., Olsen, B.B., Melau, M., Gluud, C., Nordentoft, M., 2014. Cognitive remediation combined with an early intervention service in first episode psychosis. Acta Psychiatr. Scand. 130 (4), 300–310. https://doi.org/10.1111/acps.12287.
Ott, C.V., Mine, H., Petersen, J.Z., Miskowiak, K.W., 2019. Relation between functional and cognitive impairments in remitted patients with bipolar disorder and suggestions for trials targeting cognition: an exploratory study. J. Affect. Disord. 257, 382–389. https://doi.org/10.1016/j.jad.2019.07.030.
Pan, A.W., Fisher, A.G., 1994. The assessment of motor and process skills of persons with psychiatric disorders. Am. J. Occup. Ther. 48 (9), 775–780. https://doi.org/10.5014/ajot.48.9.775.
Parsons, T.D., 2015. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. Front. Hum. Neurosci. 9, 660. https://doi.org/10.3389/fnhum.2015.00660.
Parsons, T.D., Carlew, A.R., Magtoto, J., Stonecipher, K., 2017. The potential of function-led virtual environments for ecologically valid measures of executive function in experimental and clinical neuropsychology. Neuropsychol. Rehabil. 27 (5), 777–807. https://doi.org/10.1080/09602011.2015.1109524.
Petersen, J.Z., Porter, R.J., Miskowiak, K.W., 2019. Clinical characteristics associated with the discrepancy between subjective and objective cognitive impairment in depression. J. Affect. Disord. 246, 763–774. https://doi.org/10.1016/j.jad.2018.12.105.

Rand, D., Basha-Abu Rukan, S., Weiss, P.L., Katz, N., 2009. Validation of the virtual MET as an assessment tool for executive functions. Neuropsychol. Rehabil. 19 (4), 583–602. https://doi.org/10.1080/09602010802469074.

Randolph, C., Tierney, M.C., Mohr, E., Chase, T.N., 1998. The repeatable Battery for the assessment of neuropsychological status (RBANS): preliminary clinical validity. J. Clin. Exp. Neuropsychol. 20 (3), 310–319. https://doi.org/10.1076/jcen.20.3.310.823.

Rock, P.L., Roiser, J.P., Riedel, W.J., Blackwell, A.D., 2014. Cognitive impairment in depression: a systematic review and meta-analysis. Psychol. Med. 44 (10), 2029–2040. https://doi.org/10.1017/S0033291713002535.

Rosa, A.R., Sanchez-Moreno, J., Martinez-Aran, A., Salamero, M., Torrent, C., Reinares, M., Comes, M., Colom, F., Van Riel, W., Ayuso-Mateos, J.L., Kapczinski, F., Vieta, E., 2007. Validity and reliability of the functioning assessment short test (FAST) in bipolar disorder. Clin. Pract. Epidemiol. Ment. Health 3, 5. https://doi.org/10.1186/1745-0179-3-5.

Rosa, A.R., Mercade, C., Sanchez-Moreno, J., Sole, B., Mar Bonnin, C.D., Torrent, C., Grande, I., Sugranyes, G., Popovic, D., Salamero, M., Kapczinski, F., Vieta, E., Martinez-Aran, A., 2013. Validity and reliability of a rating scale on subjective cognitive deficits in bipolar disorder (COBRA). J. Affect. Disord. 150 (1), 29–36. https://doi.org/10.1016/j.jad.2013.02.022.

Samamé, C., Durante, P., Cattaneo, B., Aprahamian, I., Strejilevich, S., 2023. Efficacy of cognitive remediation in bipolar disorder: systematic review and meta-analysis of randomized controlled trials. Psychol. Med. 53 (12), 5361–5373. https://doi.org/10.1017/s0033291723001897.

Schmidt, M., 1996. Rey auditory verbal learning test: a handbook.

Spooner, D.M., Pachana, N.A., 2006. Ecological validity in neuropsychological assessment: a case for greater consideration in research with neurologically intact populations. Arch. Clin. Neuropsychol. 21 (4), 327–337. https://doi.org/10.1016/j.acn.2006.04.004.

Torralva, T., Strejilevich, S., Gleichgerrcht, E., Roca, M., Martino, D., Cetkovich, M., Manes, F., 2012. Deficits in tasks of executive functioning that mimic real-life scenarios in bipolar disorder. Bipolar Disord. 14, 118–125. https://doi.org/10.1111/j.1399-5618.2012.00987.x.

Träger, C., Decker, L., Wæhrens, E.E., Knorr, U., Miskowiak, K., Vinberg, M., 2017. Influences of patient informed cognitive complaints on activities of daily living in patients with bipolar disorder. An exploratory cross-sectional study. Psychiatry Res. 249, 268–274. https://doi.org/10.1016/j.psychres.2016.12.058.

Tsapekos, D., Seccomandi, B., Mantingh, T., Cella, M., Wykes, T., Young, A.H., 2020. Cognitive enhancement interventions for people with bipolar disorder: a systematic review of methodological quality, treatment approaches, and outcomes. Bipolar Disord. 22 (3), 216–230. https://doi.org/10.1111/bdi.12848.

Tse, S., Chan, S., Ng, K.L., Yatham, L.N., 2014. Meta-analysis of predictors of favorable employment outcomes among individuals with bipolar disorder. Bipolar Disord. 16 (3), 217–229. https://doi.org/10.1111/bdi.12148.

Van der Elst, W., Van Boxtel, M.P., Van Breukelen, G.J., Jolles, J., 2008. A large-scale cross-sectional and longitudinal study into the ecological validity of neuropsychological test measures in neurologically intact people. Arch. Clin. Neuropsychol. 23 (7–8), 787–800. https://doi.org/10.1016/j.acn.2008.09.002.

Vandamme, D., 2010. Assessment of motor and process skills: assessing client work performance in Belgium. Work 35 (2), 201–208. https://doi.org/10.3233/wor-2010-0972.

Vita, A., Barlati, S., Ceraso, A., Nibbio, G., Ariu, C., Deste, G., Wykes, T., 2021. Effectiveness, core elements, and moderators of response of cognitive remediation for schizophrenia: a systematic review and meta-analysis of randomized clinical trials. JAMA Psychiatry 78 (8), 848–858. https://doi.org/10.1001/jamapsychiatry.2021.0620.

Voinescu, A., Petrini, K., Stanton Fraser, D., Lazarovicz, R.-A., Papavă, I., Fodor, L.A., David, D., 2023. The effectiveness of a virtual reality attention task to predict depression and anxiety in comparison with current clinical measures. Virtual Reality 27 (1), 119–140. https://doi.org/10.1007/s10055-021-00520-7.

Wæhrens, E.E., Bliddal, H., Danneskiold-Samsøe, B., Lund, H., Fisher, A.G., 2012. Differences between questionnaire- and interview-based measures of activities of daily living (ADL) ability and their association with observed ADL ability in women with rheumatoid arthritis, knee osteoarthritis, and fibromyalgia. Scand. J. Rheumatol. 41 (2), 95–102. https://doi.org/10.3109/03009742.2011.632380.

Wechsler, D., 1997. WAIS-III: Administration and Scoring Manual: Wechsler Adult Intelligence Scale. Psychological Corporation.

Wing, J.K., Tf, B., Brugha, T.T., Burke, J., Cooper, J.E., Giel, R., Jablenski, A., Regier, D., Sartorius, N., 1990. SCAN. Schedules for clinical assessment in neuropsychiatry. Arch. Gen. Psychiatry 47, 589–593. https://doi.org/10.1001/archpsyc.1990.01810180089012.

Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A., 1978. A rating scale for mania: reliability, validity and sensitivity. Br. J. Psychiatry 133, 429–435. https://doi.org/10.1192/bjp.133.5.429.