

Running title: Validation of the VRFCAT in Spain

Validation of the Virtual Reality Functional Capacity Assessment Tool (VRFCAT)

in a Spanish Sample of Individuals with Psychosis and Healthy Controls

César González-Blanch^{1,2,3}, Raquel López-Carrilero^{4,5,6}, Teresa Bobes-Bascarán^{6,7,8,9,10}, Ana Catalán^{6,11,12,13,14}, Claudia Aymerich^{6,12,15}, Roberto Rodríguez-Jiménez^{6,16,17}, Alfonso Gutiérrez-Zotes^{6,18,19,20}, Olimpia Díaz-Mandado²¹, Ana González-Pinto²², Guillermo Cano-Escalera²², Marta Zubia²², Daniel Guinart^{23,24,25}, Manuel del-Castillo-Serrano^{6,16,17}, Elisabet Vilella^{6,18,19,20}, Leticia García-Álvarez^{6,7,9,10}, Susana Ochoa^{4,5,6}, Patricia Correa-Ghisays^{6,26,27}, Vicent Balanzá-Martínez^{6,26,28,44}, Joan Vicent Sánchez-Ortí^{6,26,28}, Esther Pousa^{6,29}, Salvador Miret^{6,30,31}, Cristina Falip³⁰, Pablo Reguera-Pozuelo^{6,32,33}, Rafael Fernández-Martínez^{34,35}, Ignacio García-Cabeza^{36,37,38}, Carlos Campos-Rodríguez^{6,24}, Josep Maria Crosas³⁹, Jesus Cobo^{39,40}, Helena Pardina-Torner^{6,24}, Salvador Perona-Garcelán⁴¹, Manuel Muñoz-Caracuel^{41,42,43}, Ángel Yorca-Ruiz², Víctor Ortiz-García de la Foz², Rosa Ayesa-Arriola^{2,6}.

¹Mental Health Centre, University Hospital Marqués de Valdecilla, Santander, Spain

²IDIVAL, Valdecilla Biomedical Research Institute, Santander, Spain

³Department of Psychology, International University of La Rioja (UNIR), Logroño, Spain

⁴Research Unit, Parc Sanitari Sant Joan de Déu, Sant Boi de Llobregat, Spain

⁵MERITT Research Group, Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain

⁶Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Instituto de Salud Carlos III (ISCIII), Madrid, Spain

⁷Department of Psychology, University of Oviedo, Oviedo, Spain

⁸Department of Psychiatry, Servicio de Salud del Principado de Asturias (SESPA), Oviedo, Spain

⁹Psychiatric Research Group, Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo, Spain

¹⁰Behavioral and Mental Disorders Neuroscience Unit, INEUROPA, Oviedo, Spain

¹¹Department of Neuroscience, University of the Basque Country (UPV/EHU), Leioa, Spain

¹²Department of Psychiatry, Basurto University Hospital, OSI Bilbao-Basurto, Bilbao, Spain

This peer-reviewed article has been accepted for publication but not yet copyedited or typeset, and so may be subject to change during the production process. The article is considered published and may be cited using its DOI.

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

- ¹³Early Psychosis: Intervention and Clinical-Detection (EPIC) Lab, Department of Psychosis Studies, King's College London, UK
- ¹⁴Psychiatry Service, Biobizkaia Health Research Institute, Barakaldo, Spain
- ¹⁵Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience (IoPPN), King's College London, UK
- ¹⁶Department of Legal Medicine, Psychiatry and Pathology. Facultad de Medicina, Universidad Complutense de Madrid (UCM), Spain
- ¹⁷Department of Psychiatry, Instituto de Investigación Sanitaria Hospital 12 de Octubre (imas12), Madrid, Spain
- ¹⁸Department of Psychiatry, Hospital Universitari Institut Pere Mata, Reus, Spain
- ¹⁹Neurociències i Salut Mental, Institut d'Investigació Sanitària Pere Virgili-CERCA, Reus, Spain
- ²⁰Department Medicina i Cirurgia, Universitat Rovira i Virgili, Tarragona, Spain
- ²¹Hospitalarias Arturo Soria Foundation, Madrid, Spain
- ²²BIORABA, Department of Psychiatry, Hospital Universitario de Álava, CIBERSAM, UPV/EHU, Vitoria, Spain
- ²³Institut de Salut Mental, Hospital del Mar, Barcelona, Spain
- ²⁴Mental Health Research Group, Hospital del Mar Research Institute, CIBERSAM, Barcelona, Spain
- ²⁵Department of Psychiatry, Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA
- ²⁶Research Group on Personal Autonomy, Dependency and Severe Mental Disorders, INCLIVA Biomedical Research Institute, Valencia, Spain
- ²⁷Department of Developmental and Educational Psychology, Faculty of Psychology and Speech and Language Therapy, University of Valencia, Valencia, Spain
- ²⁸Teaching Unit of Psychiatry and Psychological Medicine, Department of Medicine, University of Valencia, Valencia, Spain
- ²⁹Department of Psychiatry, Hospital de la Santa Creu i Sant Pau, Institut de Recerca Sant Pau (IR SANT PAU), Barcelona, Spain
- ³⁰Department of Psychiatry, Mental Health and Addictions, Hospital Universitari Santa Maria, Lleida, Spain
- ³¹Biological Foundations of Mental Disorders Group, Institut de Recerca Biomèdica (IRB) de Lleida, Lleida, Spain
- ³²Department of Psychiatry, School of Medicine, University of Seville, Seville, Spain
- ³³Translational Psychiatry Group, Seville Biomedical Research Institute (IBiS)-CSIC, Seville, Spain
- ³⁴Department of Psychiatry, Hospital Álvaro Cunqueiro, Servicio Galego de Saúde (SERGAS), Vigo, Spain
- ³⁵Translational Neuroscience Group, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, Vigo, Spain
- ³⁶Department of Psychiatry, HGU Gregorio Marañón, Madrid, Spain
- ³⁷Department of Psychiatry, Hospital Dr. Rodríguez Lafora, Madrid, Spain
- ³⁸Department of Psychiatry, Universidad Francisco de Vitoria, Madrid, Spain
- ³⁹Mental Health Group, Consorci Corporació Sanitària Parc Taulí – I3PT – CIBERSAM, Sabadell, Barcelona, Spain
- ⁴⁰Departament de Psiquiatria i Medicina Legal, Universitat Autònoma de Barcelona, Barcelona, Spain
- ⁴¹Department of Psychiatry, University Hospital Virgen del Rocío, Andalusian Health Service, Seville, Spain
- ⁴²Translational Psychiatry Group, Seville Biomedical Research, Seville, Spain

⁴³Institute (IBiS)-CSIC, Foundation for Health Research Management in Seville, Seville, Spain

⁴⁴Department of Psychiatry, Hospital Clínic Universitari, Valencia, Spain.

Abstract

Introduction: The Virtual Reality Functional Capacity Assessment Tool (VRFCAT) is a computerized, performance-based measure developed to assess functional capacity through realistic simulations of daily activities. This study examined its psychometric properties in a Spanish sample including individuals with first-episode psychosis (FEP), schizophrenia and healthy controls.

Methods: A total of 370 participants (99 FEP, 116 schizophrenia, and 155 controls) completed the VRFCAT in a multicenter study. Internal consistency (McDonald's omega), discriminative validity (group comparisons and ROC curves), and convergent validity via correlations with cognitive performance and clinical symptoms were examined. Reference percentiles were calculated from the healthy control sample using quantile regression, stratified by age and education.

Results: Item-level VRFCAT completion times showed acceptable to good internal consistency overall and in controls and schizophrenia samples, but poor in FEP. Significant differences in VRFCAT performance emerged ($\chi^2 = 108.88$, $p < .001$), with controls performing best, schizophrenia worst, and FEP in between. ROC analyses indicated good discriminative accuracy in distinguishing patients from controls (area under the curve [AUC] = 0.779, sensitivity = 80.0%, specificity = 64.2%); but limited discrimination between schizophrenia and FEP. Age and education, but not sex, significantly affected performance. VRFCAT total score showed small-to-moderate correlations with cognitive performance, and no significant associations with symptom severity.

Conclusions: The VRFCAT is a reliable and valid tool for assessing functional capacity in Spanish-speaking individuals with psychotic disorders. Its ecological validity, objectivity, adequate psychometric properties, brief administration time, and ease of use support its potential use in clinical and research settings for evaluating functional recovery and treatment outcomes.

Keywords: first-episode psychosis, schizophrenia, functional capacity, VRFCAT, validation, performance-based assessment.

Introduction

Functional impairment is a core feature of psychotic disorders and a major determinant of quality of life [1]. Growing interest has focused on the assessment of functional capacity, defined as the ability to carry out everyday tasks under standardized conditions, as a distinct and valid construct that provides a direct and ecologically meaningful estimate of disability [2]. Functional capacity is more proximal to cognitive functioning than real-world behavior, and serves as a valuable intermediate outcome in clinical trials, particularly for assessing response to interventions before changes are observed in actual functioning [3-5]. Compared to traditional assessments based on self-report or informant ratings, performance-based measures offer a more ecologically valid and objective means of evaluating an individual's functional potential, while minimizing the influence of insight, environment, or support systems [6]. These tools have become increasingly central in schizophrenia research, particularly in efforts to define and measure functional recovery, where they are often used as co-primary outcomes to evaluate the impact of cognitive interventions.

The Virtual Reality Functional Capacity Assessment Tool (VRFCAT) is a computerized measure of functional capacity through realistic simulations of everyday activities [7].

Unlike traditional role-play tasks or paper-based instruments, the VRFCAT presents a series of interactive scenarios—such as meal preparation, using transportation, shopping, and handling money—designed to capture essential skills for independent living in a dynamic and ecologically valid environment. Its digital format enhances standardization, reduces administrator burden, and minimizes variability across sites, making it particularly suitable for research and clinical purposes. Importantly, the VRFCAT has demonstrated strong psychometric properties, including good test-retest reliability, sensitivity to cognitive and functional deficits, and convergent validity with both cognitive performance and real-world outcomes [7, 8]. However, key limitations remain: internal consistency—critical for interpreting the global score—has not been formally reported, and no validation studies or reference values based on healthy control samples are currently available for Spanish-speaking populations.

This study aimed to validate the VRFCAT in a Spanish sample of individuals with psychosis (schizophrenia and FEP) and healthy controls by: (i) examining internal consistency; (ii) assessing discriminative and convergent validity—expecting moderate correlations with cognitive performance and premorbid IQ, and low or non-significant correlations with symptom severity; and (iii) providing reference percentiles from the healthy control sample, stratified by sex, age, and education, to support clinical use. Based on previous literature, we hypothesized that the VRFCAT would demonstrate acceptable internal consistency and validity in a Spanish sample. Specifically, we expected patients with psychosis (schizophrenia and FEP) to perform significantly worse than healthy controls, and VRFCAT performance to show moderate associations with cognitive functioning but minimal or no associations with symptom severity.

Method

Participants and Procedure

Participants were recruited as part of a multicenter study conducted in 15 public mental health centers and university hospitals across Spain, aimed at validating digital tools for the assessment of cognitive functioning and everyday functioning in individuals with psychotic disorders (Project Code: PI20/00066). The final sample included individuals diagnosed with chronic schizophrenia (hereafter referred to as schizophrenia) or FEP, identified through clinician referral based on convenience sampling, and healthy controls selected to achieve a comparable distribution of age, sex, and educational level.

A total of 403 participants were initially recruited, including 117 individuals with FEP, 125 with schizophrenia, and 161 healthy controls. After excluding participants with missing VRFCAT data, the final sample included 370 individuals: 99 with FEP, 116 with schizophrenia, and 155 controls.

Participants were assigned to one of three groups: FEP, schizophrenia, or controls. Patients were eligible if they: (1) were aged 18–60 years, (2) had a DSM-5 diagnosis of a schizophrenia-spectrum psychotic disorder, (3) had adequate Spanish proficiency, and (4) could provide informed consent. The exclusion criteria were as follows: (1) organic brain pathology or neurological illness, (2) intellectual disability (DSM-5), and (3) current or recent (past 6 months) substance dependence, assessed via the Comprehensive Assessment of Symptoms and History [CASH; 9]. Patients were classified as FEP if they had initiated antipsychotic treatment within the past three years; otherwise, they were classified as having chronic schizophrenia.

Healthy controls met the same age and language criteria and were required to provide informed consent. Exclusion criteria included any past or current mental or neurological disorder, intellectual disability, substance use disorder, or psychotropic medication use.

Data were collected between July 2022 and December 2024. Trained research staff administered the assessments in a single session following standardized instructions. The full protocol lasted approximately 90 minutes. Written informed consent was obtained from all participants prior to the evaluation, and the study received ethical approval from the ethics committee of the principal center (code: PI20/00066) and from the corresponding ethics committees at each participating site.

Measures

Virtual Reality Functional Capacity Assessment Tool [VRFCAT; 7]

The VRFCAT is a computerized performance-based measure developed to assess functional capacity by simulating everyday tasks in a realistic virtual environment (Keefe et al., 2016). The tool presents participants with a sequence of 12 objectives embedded in a storyboard format that reflects four core domains of daily functioning: meal planning, transportation use, shopping, and money management. The Spanish version of the VRFCAT was administered on a tablet with interactive first-person navigation. The VRFCAT provides three main outcome measures: total time to completion, total number of errors, and number of forced progressions. Forced progressions occur when a participant is unable to complete a task within the maximum time limit of 300 s, in which case the system automatically registers the maximum time and advances to the next objective. The original validation study of the VRFCAT [7] found strong overlap between total time to completion and total number of errors ($r \approx$

.70). Given its greater sensitivity and reduced ceiling effects compared to total number of errors, total time to completion was selected as the primary outcome measure. A higher total time to completion indicates poorer functional capacity.

Brief Assessment of Cognition - App version [BAC App; 10]

The BAC App is a tablet-based version of the Brief Assessment of Cognition in Schizophrenia [BACS; 11], designed to assess six cognitive domains relevant to clinical populations: episodic memory, working memory, verbal fluency, processing speed, executive functioning, and psychomotor speed. The full battery takes approximately 30 minutes to complete. In the current study, we used raw scores from the BAC App. A recent validation using the same Spanish-speaking sample as the present study reported good internal consistency ($\alpha = .76-.87$), moderate positive correlations between the BAC App composite score and estimated IQ (up to $r = .49$), and good discriminative validity in distinguishing clinical from non-clinical participants [12].

Positive and Negative Syndrome Scale [PANSS; 13]

The PANSS includes three subscales: Positive Symptoms (seven items), Negative Symptoms (seven items), and General Psychopathology (16 items). In the present study, we used the validated Spanish version of the PANSS [14].

Vocabulary subtest of the Wechsler Adult Intelligence Scale – Third Edition [WAIS-III; 15]

The Spanish version Vocabulary subtest from the WAIS-III was used to estimate the premorbid intelligence quotient (IQ). Raw scores were converted to age-adjusted scaled scores using normative data from the WAIS-III manual [15].

Sociodemographic and clinical variables

Sociodemographic data, including age, sex, ethnicity, educational level, and current employment status, were collected. Clinical history was obtained through a structured interview that included mental diagnosis based on DSM-5 criteria, age of onset of duration of untreated illness (DUI), and duration of untreated psychosis (DUP).

Data Analysis

Descriptive statistics for each VRFCAT item and the total score, including means, medians, standard deviations, interquartile ranges (IQR), skewness, and kurtosis were performed separately for individuals with FEP, schizophrenia, and healthy controls. Due to the non-normal distribution of VRFCAT data, nonparametric tests were used. The internal consistency of the VRFCAT was assessed using McDonald's omega coefficient (ω), which is more appropriate for ordinal data and heterogeneous item loadings. To examine the impact of sociodemographic variables (sex, age, and years of education) on VRFCAT performance, we conducted Mann–Whitney U tests and Spearman's rank-order correlations within each group.

Quantile regression analyses were conducted on the healthy control sample to generate reference percentiles (10th, 25th, 50th, 75th, and 90th) for VRFCAT total time to completion, stratified by age and educational level. Age was categorized into three groups: 18–29, 30–44, and 45 years or older; education was grouped into 5–9, 10–14, and 15 years or more of formal schooling. These categories were used to ensure robust percentile estimation and preserve clinical interpretability. Between-group comparisons (FEP, schizophrenia, and healthy controls) were conducted using Kruskal–Wallis tests, followed by Dunn's post hoc pairwise comparisons with p-values adjusted using

Holm's correction for multiple testing. Discriminative validity was assessed using Receiver Operating Characteristic (ROC) curve analyses, reporting the area under the curve (AUC), sensitivity, specificity, and the Youden Index, which reflects the maximum combined value of sensitivity and specificity.

Finally, convergent validity was examined using Spearman's rank-order correlations between VRFCAT total time to completion and cognitive performance (BAC App) as well as estimated premorbid IQ (WAIS-IV Vocabulary), across the three groups. Associations with clinical symptom severity (PANSS) were also explored to examine discriminant validity.

All statistical tests were two-tailed, with the significance threshold set at $p < .05$. Analyses were conducted using IBM SPSS Statistics (version 26) and JASP (version 0.19.3).

Results

Sociodemographic and Clinical Characteristics of the Sample

Table 1 shows demographic and clinical characteristics across groups. Significant differences were found across groups in age, years of education, ethnicity, marital status, and employment status. In contrast, sex distribution did not differ significantly between groups.

We compared sex, age, and education between the clinical ($n = 215$) and control ($n = 155$) groups to assess demographic comparability. No significant differences were found in sex ($\chi^2(1) = 2.75, p = .097$) or age ($U = 15,455.5, p = .628$). Although years of

education differed significantly ($U = 12,257.5$, $p < .001$), the nonparametric effect size was modest ($r = .23$).

Table 2 presents descriptive statistics for each VRFCAT item and the total time to completion across the three groups. The distribution showed marked deviations from normality, with several items exhibiting substantial positive skewness and kurtosis.

Internal Consistency

Internal consistency was acceptable to good across groups: for the total sample, $\omega = 0.80$ (95% CI [0.77–0.83]); for the clinical group, $\omega = 0.76$ (95% CI [0.71–0.81]); and for controls, $\omega = 0.71$ (95% CI [0.64–0.78]). When FEP and schizophrenia groups were examined separately, internal consistency was acceptable in the schizophrenia ($\omega = 0.77$, 95% CI [0.71–0.83]), but low in FEP ($\omega = 0.34$, 95% CI [0.20–0.49]).

Effects of Sex, Age, and Education on VRFCAT Performance

No sex differences in VRFCAT performance were observed within any group (all $p > 0.05$). In contrast, performance was significantly associated with age and education in some groups. In the FEP group, neither age nor education showed significant associations with VRFCAT performance (age: $\rho = .165$, $p = .106$; education: $\rho = -.167$, $p = .099$). In the schizophrenia group, older age was moderately correlated with slower VRFCAT performance ($\rho = .408$, $p < .001$), and fewer years of education were associated with poorer performance ($\rho = -.221$, $p = .017$). When combining the FEP and schizophrenia groups, age and education were significantly associated with VRFCAT total score (age: $\rho = .403$, $p < .001$; education: $\rho = -.243$, $p < .001$). In healthy controls, associations were weaker: VRFCAT showed a modest negative correlation with years of

education ($\rho = -.158$, $p = .050$), and a positive but non-significant correlation with age ($\rho = .150$, $p = .067$). Based on these findings, percentile reference values were derived from the healthy control group according to age and education, but not by sex.

Reference Values from the Healthy Control Sample

Reference values for VRFCAT total time to completion were calculated using quantile regression in the control sample, stratified by age group and educational level.

Percentiles (10th, 25th, 50th, 75th, and 90th) were generated to enhance the clinical utility and interpretability of the measure (Table 3).

Discriminative Validity

ROC analyses were conducted to assess the ability of the VRFCAT total time to completion to distinguish between groups. The AUC for differentiating the clinical group (schizophrenia and FEP combined) from healthy controls was 0.779 (95% CI [0.732, 0.826]), indicating good accuracy. The optimal cutoff point (732 s) was identified based on the Youden Index (0.44), which yielded a sensitivity of 80.0% and specificity of 64.2%. When comparing schizophrenia and FEP, the AUC was 0.721 (95% CI [0.653, 0.788]), indicating modest discriminative power. The optimal cutoff point (746 s), based on the Youden Index (0.28), yielded a sensitivity of 74.1%, specificity of 53.5%.

Group Differences in Item-Level Performance

Table 4 presents group differences in VRFCAT. Significant effects were found for all items and for the total time to completion (all $p < .05$). *Post hoc* comparisons revealed that participants with schizophrenia consistently had the longest completion times,

performing significantly worse than both healthy controls and individuals with FEP on all items except item 4, for which no significant differences were found between schizophrenia and controls. This pattern was further confirmed by the total score, which also differed significantly across groups ($\chi^2 = 108.9$, $p < .001$): the schizophrenia group showed the poorest performance, controls the best, and the FEP group fell in between.

Associations with Cognitive Performance, IQ, and Clinical Symptoms

As shown in Table 5, VRFCAT total time to completion was significantly correlated with several cognitive measures across all groups, with small to moderate associations, strongest in the schizophrenia group. No significant association was observed with symptom severity in either clinical group.

Discussion

The aim of the present study was to examine the psychometric properties of the VRFCAT in a Spanish sample of individuals with FEP and schizophrenia and controls and to provide reference values from the control group to support the interpretation of performance in individuals with psychosis. First, item-level completion times on the VRFCAT demonstrated acceptable to good internal consistency across the total sample and groups. Second, the tool showed adequate discriminative validity, effectively distinguishing between groups, with controls performing best, schizophrenia the worst, and FEP at an intermediate level. ROC curve analyses further supported the discriminative ability of the VRFCAT, revealing good accuracy in differentiating the clinical group from healthy controls, but only modest performance in distinguishing between FEP and schizophrenia. Third, convergent validity was supported by modest correlations between VRFCAT performance and cognitive functioning. In contrast, no

significant associations were found with symptom severity, consistent with the expected discriminant validity. Finally, we provided stratified reference percentiles from the healthy control sample by age and education to aid interpretation of scores in clinical and research contexts.

The current findings are broadly consistent with those of previous studies, validating the VRFCAT as a sensitive tool for assessing functional capacity in individuals with psychosis. Across several studies, the VRFCAT has consistently demonstrated sensitivity to group differences, with individuals with schizophrenia or FEP performing significantly worse than healthy controls [7, 8, 16, 17]. Importantly, its ability to differentiate between individuals with schizophrenia and those with FEP was modest. This may reflect the early emergence of functional deficits and heterogeneity both within and across phases of psychosis, as well as the nature of the VRFCAT as a measure of potential functional ability under optimal conditions, which may not fully capture the real-world functional deterioration typically observed in chronic stages.

In terms of reliability, the test–retest reliability of the VRFCAT has been previously demonstrated in both clinical and non-clinical populations [7, 18], however, to our knowledge, internal consistency has not been assessed to date. In our study, we observed acceptable to good reliability across groups, as indicated by McDonald's omega coefficients, supporting the robustness and reliability of the total score as a performance-based outcome. However, reliability was notably lower in the FEP subgroup, suggesting that measurement consistency may be affected in the early stages of illness, possibly due to greater clinical heterogeneity, and highlighting the need for further investigation in this population.

Although few studies have examined the influence of sociodemographic variables on VRFCAT performance [7, 8, 18, 19], our findings are consistent with the limited evidence available. Older participants and those with lower educational attainment had slower completion times. This pattern aligns with the findings of Atkins et al. [18], who used the VRFCAT in a healthy sample and reported significant age-related differences, likely reflecting declines in processing speed, working memory, and executive functioning, which impact the performance of complex daily tasks. Additionally, Ventura et al. [8] found that while patient–control differences in VRFCAT total time to completion remained significant after adjusting for parental education (used as a proxy for expected educational level), the effect was attenuated, supporting the role of educational background in influencing functional capacity performance, likely through its impact on the cognitive and practical skills needed for everyday tasks assessed by the VRFCAT. In contrast, no significant sex differences were found in our sample, neither in patients nor in controls, a finding that is in line with previous studies using this tool [7, 8, 19].

Slower VRFCAT performance was significantly associated with poorer cognitive functioning. However, the small to moderate effect sizes of these correlations suggest that these constructs are related but distinct. Correlations with neurocognitive performance assessed using the MATRICS Consensus Cognitive Battery (MCCB) have been consistently reported in previous studies [7, 8, 17, 20]. Harvey et al. [20] conducted a joint factor analysis of cognitive and functional capacity measures and found that, while the relationship between the MCCB and the UCSD Performance-Based Skills Assessment (UPSA)—a widely used measure of functional capacity— fit a one-factor model, whereas the VRFCAT required a two-factor solution, with the factors

being moderately correlated. These findings support the notion that the VRFCAT captures functionally relevant abilities that extend beyond those measured by traditional neurocognitive assessments. Our results should also be considered in light of current international recommendations, such as the recent European Psychiatric Association guidance on the treatment of cognitive impairment in schizophrenia, which emphasizes cognition as a key therapeutic target to improve functional capacity and, ultimately, daily functioning [21].

This study has several limitations. First, the cross-sectional design precludes any conclusions about the temporal stability of VRFCAT performance or its predictive value for real-world functional outcomes. Second, convergent validity was examined only through associations with few cognitive measures (BAC App and IQ estimation), and no additional functional capacity or real-world functioning assessments were included for comparison. In particular, the lack of direct comparison with other validated performance-based measures represents a key limitation for assessing convergent validity; this should be addressed in future studies through concurrent administration of the VRFCAT and the UPISA-Brief [22] or similar instruments. Third, participants were recruited through convenience sampling. Third, internal consistency for the VRFCAT in the FEP subgroup was unacceptably low, raising concerns about the reliability of the measure in early psychosis. Finally, the sample size was determined based on psychometric considerations derived from the Rasch model [23] rather than on a conventional a priori power analysis. Although the final sample (215 patients and 155 controls) did not reach the originally intended size (≈ 250 per group) recommended for maximum calibration stability, it was sufficient for the planned validation analyses.

These findings lead to several recommendations for future research using the VRFCAT.

First, given the non-normal distribution typically observed in task completion times, non-parametric statistical methods are strongly recommended, especially with small samples ($N \leq 30$) or strong deviations from normality. Second, given that total time to completion is frequently used as the primary outcome in VRFCAT studies, it is surprising that its internal consistency has not been previously reported; this metric should be routinely included in future research. Third, participant age and educational attainment should be considered in both study design and interpretation of results, as these variables have consistently shown to influence VRFCAT performance.

Additionally, although no significant sex differences have been observed to date, continued evaluation of potential gender-related effects is warranted to ensure that conclusions drawn from VRFCAT data are generalizable to both men and women.

The present study has several strengths worth noting. First, it provides the first validation of the VRFCAT in a Spanish-speaking sample, addressing an important gap in the international literature and expanding the potential for cross-cultural research and clinical application. Moreover, the study was conducted across multiple centers, enhancing the generalizability of the findings within the Spanish clinical context.

Interestingly, previous international research has shown that regional differences in performance on functional capacity assessments such as the VRFCAT are modest compared to the robust and consistent differences observed between individuals with schizophrenia and healthy controls [16]. This supports that, despite potential cultural and geographical variations, the pattern of functional impairment captured by the VRFCAT is likely to be generalizable across settings, supporting its utility for both clinical and research purposes beyond the Spanish population. Second, the inclusion of

individuals with schizophrenia, FEP and demographically comparable controls allows for a more detailed understanding of functional capacity across different stages of illness. Third, the use of quantile regression to generate stratified reference values from the healthy control sample by age and education enhances the clinical interpretability of individual scores.

In summary, the present findings support the VRFCAT as a reliable and valid tool for assessing functional capacity in individuals with psychosis within a Spanish-speaking population. The instrument demonstrated adequate psychometric properties, including internal consistency, discriminative ability, and convergent validity with cognitive functioning. Its performance-based nature, ecological validity, and standardized format make it particularly valuable for both research and clinical contexts, especially given its brief administration time, ease of use, and minimal demands on staff. The VRFCAT represents a promising measure for assessing functional recovery and may serve as a useful endpoint in intervention studies targeting everyday functioning in psychosis.

Declarations

Funding

This study was supported by the Instituto de Salud Carlos III (ISCIII) through project PI20/00066, co-funded by the European Union.

Ana González-Pinto acknowledges the support of the Spanish Ministry of Science, Innovation and Universities, integrated into the National R&D&I Plan and co-funded by the ISCIII–Subdirección General de Evaluación and the European Regional Development Fund (FEDER; grants PI18/01055 and PI21/00713), as well as the CIBERSAM, the Basque Government (grant 2022111054), and the University of the Basque Country (grant IT1631-22).

Ángel Yorca-Ruiz is funded by a predoctoral contract (PFIS FI/00162) from the Carlos III Health Institute. Rosa Ayesa-Arriola was funded by a Miguel Servet contract from the Carlos III Health Institute (CP18/00003) and a Consolidator Grant from the Spanish Ministry of Science and Innovation (CNS2022-136110).

Acknowledgements

We would like to extend our sincere gratitude to all the participants who generously contributed their time and effort to this research. We are especially grateful to the 15 collaborating hospitals and research centers across Spain whose support and collaboration made this work possible. The dedication of the clinical and research teams at each site was essential to the successful execution of this multicenter study. We also thank the developers of the Virtual Reality Functional Capacity Assessment Tool (VRFCAT) for their support and assistance during the validation process.

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of the Hospital Universitario Marqués de Valdecilla (Approval Code: PI20/00066) and from the corresponding ethics committees of each participating site. Written informed consent was obtained from all participants prior to study enrollment.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions (CRediT statement)

CGB: Data curation, Formal analysis, Visualization, Writing – Original Draft, Writing – Review & Editing.

VOG: Data curation, Writing – Review & Editing

RAA: Conceptualization, Methodology, Funding acquisition, Project administration, Writing – Review & Editing.

RLC, TBB, AC, CA, RRJ, AGZ, ODM, AGP, GCE, MZ, DG, MdCS, EV, LGA, SO, PCG, VBM, JVSO, EP, SM, CF, PRP, RFM, IGC, CCR, JMC, JC, AYR: Investigation, Writing – Review & Editing.

References

- [1] Nevarez-Flores AG, Sanderson K, Breslin M, Carr VJ, Morgan VA, Neil AL. Systematic review of global functioning and quality of life in people with psychotic disorders. *Epidemiol Psychiatr Sci.* 2019;28:31-44.
- [2] Harvey PD, Velligan DI, Bellack AS. Performance-based measures of functional skills: usefulness in clinical treatment studies. *Schizophr Bull.* 2007;33:1138-48.
- [3] Czaja SJ, Loewenstein DA, Lee CC, Fu SH, Harvey PD. Assessing functional performance using computer-based simulations of everyday activities. *Schizophr Res.* 2017;183:130-6.
- [4] Harvey PD, Chirino M, Mueller A, Rivera-Molina A, Zayas-Bazan M, Kallestrup P, et al. Improvements in performance based measures of functional capacity and cognition after computerized functional skills training in older people with mild cognitive impairment and healthy comparators. *Psychiatry Res.* 2024;334:115792.
- [5] Green MF, Nuechterlein KH, Kern RS, Baade LE, Fenton WS, Gold JM, et al. Functional co-primary measures for clinical trials in schizophrenia: results from the MATRICS Psychometric and Standardization Study. *Am J Psychiatry.* 2008;165:221-8.
- [6] Mantovani LM, Teixeira AL, Salgado JV. Functional capacity: a new framework for the assessment of everyday functioning in schizophrenia. *Braz J Psychiatry.* 2015;37:249-55.
- [7] Keefe RS, Davis VG, Atkins AS, Vaughan A, Patterson T, Narasimhan M, et al. Validation of a Computerized test of Functional Capacity. *Schizophr Res.* 2016;175:90-6.
- [8] Ventura J, Welikson T, Ered A, Subotnik KL, Keefe RSE, Helleman GS, et al. Virtual reality assessment of functional capacity in the early course of schizophrenia: Associations with cognitive performance and daily functioning. *Early Interv Psychiatry.* 2020;14:106-14.
- [9] Andreasen NC, Flaum M, Arndt S. The Comprehensive Assessment of Symptoms and History (CASH). An instrument for assessing diagnosis and psychopathology. *Arch Gen Psychiatry.* 1992;49:615-23.
- [10] Atkins AS, Tseng T, Vaughan A, Twamley EW, Harvey P, Patterson T, et al. Validation of the tablet-administered Brief Assessment of Cognition (BAC App). *Schizophr Res.* 2017;181:100-6.
- [11] Keefe RS, Goldberg TE, Harvey PD, Gold JM, Poe MP, Coughenour L. The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophr Res.* 2004;68:283-97.
- [12] González-Blanch C, Pousa E, Reguera-Pozuelo P, Gutiérrez-Zotes A, Rodríguez-Jiménez R, del-Castillo-Serrano M, et al. Validation and Normative Data for the BAC App in Spanish-Speaking Individuals with Psychosis and Healthy Controls. *Eur Psychiatry.* in press.
- [13] Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull.* 1987;13:261-76.
- [14] Peralta V, Cuesta MJ. [Validation of positive and negative symptom scale (PANSS) in a sample of Spanish schizophrenic patients]. *Actas Luso Esp Neurol Psiquiatr Cienc Afines.* 1994;22:171-7.
- [15] Wechsler D. WAIS-III. Escala de inteligencia de Wechsler para adultos-III [Wechsler Adult Intelligence Scale-III]. Madrid: TEA Ediciones; 1999.
- [16] Guzman A, Klein H, Keefe RS, Horan WP, Harvey PD. Location-based differences in cognition and functional capacity: Consistent levels of impairment in participants with schizophrenia compared to healthy controls. *Psychiatry Res.* 2024;340:116121.
- [17] Ruse SA, Harvey PD, Davis VG, Atkins AS, Fox KH, Keefe RS. Virtual Reality Functional Capacity Assessment In Schizophrenia: Preliminary Data Regarding Feasibility and Correlations with Cognitive and Functional Capacity Performance. *Schizophr Res Cogn.* 2014;1:e21-e6.
- [18] Atkins AS, Stroescu I, Spagnola NB, Davis VG, Patterson TD, Narasimhan M, et al. Assessment of Age-Related Differences in Functional Capacity Using the Virtual Reality Functional Capacity Assessment Tool (VRFCAT). *J Prev Alzheimers Dis.* 2015;2:121-7.

- [19] Harvey PD, Keefe RSE, Kallestrup P, Czaja SJ, Klein H, Horan W. Migration of digital functional capacity assessments from device resident to cloud-based delivery: Development and convergent validity. *Schizophr Res Cogn*. 2025;39:100331.
- [20] Harvey PD, Horan WP, Atkins AS, Stevens H, Welch M, Yuan J, et al. Factor structure of cognitive performance and functional capacity in schizophrenia: Evidence for differences across functional capacity measures. *Schizophr Res*. 2020;223:297-304.
- [21] Vita A, Gaebel W, Mucci A, Sachs G, Erfurth A, Barlati S, et al. European Psychiatric Association guidance on assessment of cognitive impairment in schizophrenia. *Eur Psychiatry*. 2022;65:e58.
- [22] Mausbach BT, Harvey PD, Goldman SR, Jeste DV, Patterson TL. Development of a brief scale of everyday functioning in persons with serious mental illness. *Schizophr Bull*. 2007;33:1364-72.
- [23] JM L. Sample size and item calibration stability. *Rasch Measurement Transactions*. 1994;7:328.

Table 1. Demographic and Clinical Characteristics of FEP, Schizophrenia, and Healthy Controls

	FEP (n = 99)	Schizophrenia (n = 116)	HC (n = 155)	Statistic	P- value
Age, Md (IQR)	24 (20.5–33)	40 (31–48)	31 (26–37)	H = 67.930	< .001
Sex, n (%)				$\chi^2 = 3.068$	.216
Male	62 (62.6)	77 (66.4)	87 (56.1)		
Female	37 (37.4)	39 (33.6)	68 (43.9)		
Years of education Md (IQR)	14.0 (11–17)	12.0 (10–16)	16 (12–19)	H = 22.343	< .001
Marital status, , n (%)				$\chi^2 = 23.047$	< .001
Single	81 (81.8)	94 (81.0)	92 (59.4)		
Married	12 (12.1.9)	13 (11.2)	47 (30.3)		
Other	5 (5.1)	9 (7.8)	12 (7.7)		
Ethnicity, n (%)				$\chi^2 = 30.657$	< .001
Caucasian	52 (52.5)	94 (81.0)	126 (81.3)		
Latino	43 (43.4)	20 (17.2)	27 (17.4)		
Other	4 (4.0)	2 (1.7)	2 (1.3%)		
Employment status, n (%)				$\chi^2 = 165.437$	< .001
Student	26 (26.3)	11 (9.5)	27 (17.4)		
Employed	36 (36.4)	21 (18.1)	112 (72.3)		
Unemployed	20 (20.2)	19 (16.4)	9 (5.8)		
Disability	2 (2.0)	48 (41.4)	2 (1.3)		
Other	13 (13.1)	17 (14.7)	5 (3.2)		
DUP, Md (IQR)	2 (1–7)	2 (1–24)	—	U = 4814.0	.078
DUI, Md (IQR)	10 (2–24)	12.0 (1–95)	—	U = 5099.0	.435
PANSS Positive, Md (IQR)	10 (7–18)	11.0 (8–16)	—	U = 5244.0	.893
PANSS Negative, Md (IQR)	11 (7–17)	15.0 (10–20)	—	U = 3905.5	.008
PANSS General, Md (IQR)	25.5 (20–35.5)	25.0 (20–31)	—	U = 4922.5	.695

Note. FEP = first-episode psychosis; HC = healthy controls; DUP = duration of untreated psychosis (in months); DUI = duration of untreated illness (in months); PANSS = Positive and Negative Syndrome Scale; Md = median; IQR = interquartile range (25th–75th percentile). H = Kruskal–Wallis test statistic; U = Mann–Whitney U test statistic.

Table 2. Descriptive Statistics for VRFCAT Items in FEP, Schizophrenia, and Healthy Control Groups

VRFCAT Item	Mean	SD	Median	IQR	Skewness	Kurtosis
<i>FEP</i>						
1	49.9	16.5	46.3	48.1 - 45.7	5.9	42.2
2	87.7	41.3	74.8	100.0 - 63.9	3.1	13.0
3	121.6	77.3	104.0	168.1 - 58.0	0.8	-0.2
4	22.8	7.0	21.2	22.0 - 21.0	7.8	69.4
5	12.7	6.0	10.7	13.1 - 9.8	3.8	17.7
6	69.3	24.7	65.6	67.2 - 65.0	8.4	80.0
7	44.3	11.5	41.6	46.1 - 38.0	1.9	5.7
8	27.4	20.5	23.0	27.7 - 19.0	6.2	45.5
9	134.8	49.5	128.0	162.0 - 102.0	0.9	1.7
10	85.8	54.0	66.0	100.7 - 49.9	1.9	4.3
11	68.2	16.5	64.5	66.1 - 64.0	5.2	42.0
12	45.4	39.8	34.0	51.0 - 24.8	3.7	18.1
Total time	769.8	193.5	722.7	855.0 - 629.7	0.6	1.2
<i>Schizophrenia</i>						
1	51.1	8.6	47.2	52.1 - 46.2	2.3	5.4
2	101.8	43.2	91.1	120.5 - 68.8	1.9	4.6
3	148.7	89.8	125.9	217.0 - 71.9	0.5	-1.1
4	26.3	22.0	21.7	22.5 - 21.0	7.3	59.5
5	18.4	16.7	12.1	18.0 - 10.6	3.8	17.5
6	79.4	42.6	66.2	73.2 - 65.2	4.4	19.3
7	63.1	44.2	49.2	67.8 - 39.3	3.6	15.5
8	31.8	12.4	29.6	38.0 - 23.2	1.6	3.1
9	194.8	70.4	183.8	269.3 - 133.9	0.3	-1.3
10	108.1	65.5	86.4	125.4 - 61.5	1.5	1.8
11	77.8	39.6	65.0	68.4 - 64.0	3.8	14.1
12	74.1	59.9	52.6	90.0 - 33.0	2.0	4.3
Total time	975.5	300.2	943.6	1088.1 - 737.8	1.2	1.9
<i>Controls</i>						
1	50.5	19.5	46.1	45.6 - 48.0	6.5	46.3
2	78.8	26.4	69.6	60.1 - 93.9	1.5	3.2
3	72.7	46.0	50.0	39.2 - 99.3	1.9	5.0
4	23.6	6.7	21.3	21.0 - 22.6	3.9	16.9
5	14.0	10.2	10.7	9.8 - 13.4	5.1	32.7
6	68.0	7.4	65.7	65.0 - 67.0	3.7	15.3
7	44.6	14.3	41.0	36.9 - 47.2	3.3	16.4
8	25.1	7.5	23.6	19.6 - 28.4	1.7	4.3
9	116.7	41.3	108.1	91.0 - 128.1	1.9	5.2
10	68.3	40.7	56.0	47.3 - 72.5	3.4	14.1
11	67.9	20.2	64.8	64.0 - 66.1	10.4	116.2
12	37.1	30.0	29.8	21.1 - 43.2	5.1	39.0
Total time	667.2	149.0	637.0	575.5 - 714.7	2.5	10.3

Note. VRFCAT = Virtual Reality Functional Capacity Assessment Tool; FEP = First-Episode Psychosis; SD = standard deviation; IQR = Interquartile Range (25th to 75th percentile).

Table 3. Estimated VRFCAT Reference Values by Age and Education Level
Derived from Healthy Controls

Age group (years)	Years of education			Percentile
	5–9	10–14	15+	
18–29	537	511	512	90
	682	597	583	75
	780	625	611	50
	809	674	672	25
	886	736	754	10
30–44	555	529	530	90
	660	575	561	75
	798	642	628	50
	845	710	708	25
	924	775	792	10
45+	570	543	544	90
	666	582	567	75
	846	690	677	50
	909	774	772	25
	978	828	845	10

Note. Values represent VRFCAT total time to completion (in seconds) estimated from quantile regression according to age group and education level. Percentiles (10th, 25th, 50th, 75th, 90th) are presented in descending order to reflect that lower rows correspond to lower performance, as higher completion times indicate poorer functional capacity. For example, a 35-year-old control participant with 12 years of education who takes 775 seconds to complete the VRFCAT would fall at the 10th percentile for their demographic group, suggesting a performance approximately one standard deviation below the expectations derived from the healthy control sample.

Table 4. Kruskal–Wallis Test Comparisons of VRFCAT Item-Level and Total Time to Completion Across FEP, Schizophrenia, and Control Groups

Item	χ^2 (df = 2)	p-value	Mean ranks			Post-hoc (Dunn) ¹
			FEP (n = 99)	SCZ (n = 116)	HC (n = 155)	
1. Pick up the recipe	22.83	< .001	173.69	224.38	163.95	SCZ > FEP (z = -3.464, p = .001) SCZ > HC (z = 4.603, p<.001)
2. Search for Ingredients	29.03	< .001	182.62	226.94	156.32	SCZ > FEP (z = -3.029, p = .005) SCZ > HC (z = 5.378, p<.001)
3. Cross off correct ingredients and pick up the bus schedule	68.19	< .001	206.49	237.42	133.23	SCZ > HC (z = 7.936, p<.001) FEP > HC (z = 5.325, p<.001)
4. Pick up the billfold	9.099	0.011	164.3	207.84	182.32	SCZ > FEP (z = -2.114, p = .035) SCZ > FEP (z = -2.977, p = .009)
5. Exit the apartment	21.76	< .001	165.21	223.78	169.82	SCZ > HC (z = 4.110, p<.001) SCZ > FEP (z = -4.003, p<.001)
6. Get on the bus to the grocery store	13.02		165.84	214.51	176.35	SCZ > FEP (z = -3.327, p = .003) SCZ > HC (z = 2.908, p = .007)
7. Pay for the bus	31.27	< .001	167.86	231.4	162.42	SCZ > HC (z = 5.253, p<.001) SCZ > FEP (z = -4.342, p<.001)
8. Select an aisle	34.08	< .001	160.48	233.43	165.61	SCZ > HC (z = 5.165, p<.001) SCZ > FEP (z = -4.985, p<.001)
9. Shop for groceries	100.48	< .001	177.92	263.28	132.13	SCZ > HC (z = 9.990, p<.001) SCZ > FEP (z = -5.834, p<.001)
10. Pay for groceries	47.30	0.013	188.09	236.16	145.93	FEP > HC (z = 3.328, p<.001) SCZ > HC (z = 6.872, p<.001)
11. Get on the bus to go home	8.69	< .001	169.46	209.2	178.01	SCZ > FEP (z = -3.285, p = .002) FEP > HC (z = 3.064, p = .002)
12. Pay for the bus	54.96	< .001	175.11	244.32	148.11	SCZ > FEP (z = -2.717, p = .020) SCZ > HC (z = 2.377, p = .035)
						SCZ > HC (z = 7.327, p<.001)
Total time	108.88	< .001	189.24	262.44	125.54	SCZ > FEP (z = -4.729, p<.001) FEP > HC (z = 1.962, p = .050) SCZ > HC (z = 10.427, p<.001) SCZ > FEP (z = -5.002, p<.001) FEP > HC (z = 4.630, p<.001)

Note: Df = degrees of freedom; SCZ = schizophrenia; FEP = first-episode psychosis; HC = healthy controls.

¹ Post hoc p-values are adjusted using the Holm correction for multiple comparisons. Group comparisons (e.g., SCZ > FEP) reflect differences in mean ranks; higher ranks correspond to longer completion times and thus poorer VRFCAT performance.

Table 5. Spearman Correlations Between VRFCAT Total Time To Completion and IQ, BAC App, and PANSS Scores in Individuals with FEP, Schizophrenia, and Healthy Controls

Variable	FEP (n = 99)		Schizophrenia (n = 116)		Healthy Controls (n = 155)	
	ρ	p-value	ρ	p-value	ρ	p-value
IQ (Vocabulary)	0.001	0.994	-0.306	0.001	-0.174	0.037
BAC App - Verbal Memory	-0.352	< 0.001	-0.387	< 0.001	-0.172	0.033
BAC App - Digit Sequencing	-0.127	0.209	-0.310	0.001	-0.185	0.021
BAC App - Token Motor Task	-0.426	< 0.001	-0.498	< 0.001	-0.225	0.005
BAC App - Fluency	-0.065	0.521	-0.247	0.008	-0.214	0.007
BAC App - Symbol Coding	-0.368	< 0.001	-0.622	< 0.001	-0.301	< 0.001
BAC App - Tower of London	-0.218	0.030	-0.572	< 0.001	-0.110	0.172
PANSS Positive	0.092	0.379	0.102	0.281	-	-
PANSS Negative	0.128	0.230	0.077	0.419	-	-
PANSS General	0.082	0.441	0.077	0.418	-	-

Note. FEP = First-Episode Psychosis; VRFCAT = Virtual Reality Functional Capacity Assessment Tool; PANSS = Positive and Negative Syndrome Scale; IQ = estimated Intelligence Quotient; BAC App = Brief Assessment of Cognition App. Bolded values indicate statistically significant correlations ($p < .05$).