Breaking Cultural Barriers: Culturally Fair Assessment of Memory and Other Cognitive Skills Within the Black Community.

Tolulope Odumuyiwa

DClinPsy thesis (Volume 1), 2025

University College London

UCL Doctorate in Clinical Psychology

Thesis declaration form

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Name: Tolulope Odumuyiwa

Date: 12/06/2025

Overview

Cognitive screening tools are central to dementia diagnosis and prevalence research. However, concerns have been raised about their cultural fairness, particularly for Black African and Caribbean populations, who are often underrepresented in dementia research and at risk of diagnostic bias. This thesis therefore sought to explore the psychometric quality and cultural relevance of cognitive screening tools used in Sub-Saharan Africa (SSA), and to examine whether culture-fair tools offer a more equitable alternative to standard assessments in a UK context.

Part 1 comprises a systematic review of cognitive screening tools used in SSA dementia prevalence studies. Twelve tools were identified from nine studies, and their psychometric properties were appraised using the Terwee et al. (2007) criteria. Whilst the Rowland Universal Dementia Assessment Scale (RUDAS) demonstrated strong psychometric and contextual performance, several tools lacked local validation. Findings emphasised the need for culturally grounded, African-led tool development and clearer guidance for clinical use.

Part 2 comprises an empirical study comparing performance on standard and culture-fair cognitive tests in a sample of cognitively healthy Black African/Caribbean and White British adults in the UK. A cross-sectional design and Bayesian analyses were employed. Between-group comparisons showed strong evidence for no difference across most tests. However, within-group analyses revealed that Black participants performed better on culture-fair tools, which were also rated as more intuitive, relevant, and comfortable.

Part 3 comprises a critical appraisal of the research process. Topics explored include: Reflections on terminology, bias and cultural fairness; researcher reflexivity and positionality; community-engaged recruitment; learning Bayesian methods; the development of practitioner identity; and reflections for future research and clinical practice.

Impact statement

This thesis addresses long-standing concerns regarding the cultural fairness of cognitive screening tools used in dementia research and diagnosis, particularly for individuals of Black African and Caribbean heritage. While dementia prevalence is rising globally, Black communities remain underrepresented in research. They are at increased risk of diagnostic inequities due to many factors, including the reliance on standardised tools developed in Western contexts. This thesis directly addresses this issue by examining the psychometric quality and cultural relevance of cognitive screening tools used in Sub-Saharan Africa (SSA), as well as performance on both standardly used and culture-fair assessments in a UK-based sample focused on ethnic disparities, specifically between Black African/Caribbean and White British participants.

The empirical study is one of the first in the UK to apply the Cross-Cultural Dementia screening test (CCD) with Black African and Black Caribbean participants. This is a novel and meaningful contribution that begins to fill a critical gap in the evidence base regarding how these tools perform across cultural contexts. The study further contributes to the relatively under-used application of Bayesian analysis in clinical psychology and neuropsychology, offering a way to model performance differences that is sensitive to uncertainty and avoids overreliance on p-values.

In practical terms, the research has already had a tangible impact by enhancing the representation of Black participants in the Dementia Research Centre's database, supporting efforts to ensure more equitable research, and indirectly informing future clinical guidance and practice. Additionally, participant feedback collected as part of the study has provided valuable insight into how culturally grounded and standard tools are experienced by minoritised ethnic individuals, an area often overlooked in neuropsychology research.

Looking ahead, the findings could inform NHS memory services and research institutions working with diverse populations. For example, the evidence that culture-fair tools such as the CCD and RUDAS are more intuitive and comfortable for Black participants supports arguments for their wider adoption in clinical screening processes, potentially reducing misdiagnosis and improving access to appropriate care. The thesis underscores the importance of contextual validation and culturally informed assessment practices, addressing

the lack of specific guidance in NICE dementia guidelines on how to adapt assessments for diverse populations.

Theoretically, this thesis challenges the assumption of universality in cognitive assessments and advocates for a more pluralistic approach to neuropsychology. It demonstrates how culture can influence test performance and highlights the need for more inclusive frameworks in both tool development and clinical training. While the study was not designed to produce a clinical tool or intervention, it contributes foundational evidence that may influence future research, guidance, and practice in culturally competent dementia care in both UK and international settings.

In summary, this thesis provides a platform for future work to advance health equity, particularly for Black communities that are often marginalised in cognitive health research. It signals a shift towards more inclusive science and practice, with the potential to impact how dementia is understood, measured, and managed in multi-ethnic contexts.

Table of Contents

Part 2

List of Tables

Part 1

Part 2

List of Figures

Part 1

Part 2

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to all the participants who took part in this study. Thank you for sharing your time, and your reflections on the process. I was moved by the many personal conversations we had, whether about loved ones affected by dementia or your desire to contribute to research that centres the Black community. Your openness and generosity gave this work its meaning and purpose.

I am also deeply thankful to the organisations that supported this project by sharing the study across your networks. Alzheimer's Society UK, Join Dementia Research, Dementia UK helped extend its reach, while community-based groups like the Caribbean Social Forum played an especially meaningful role. I'm particularly grateful to Pamela from the Caribbean Social Forum, who generously took the time to speak with me about meaningful community engagement and shared the vital work being done within the Forum to foster culturally inclusive environments for people affected by dementia. I look forward to continuing those conversations through future workshops. To members of The Redeemed Christian Church of God, Living Water Parish, particularly Pastor Modupe Afolabi, thank you for so warmly welcoming me into your congregation. Your generosity in supporting recruitment and welcoming me to run dementia workshops in your church had a lasting impact, creating space for much-needed dialogue.

To my external supervisor, Dr Jess Jiang, thank you for your invaluable input throughout the development of this project. Your guidance on the study's design, statistics, and the practical support provided through the Dementia Research Centre, made this work possible. I have learned a great deal from your thoughtful direction.

To my internal supervisor, Dr Georgina Charlesworth, thank you for encouraging me to pursue this project when it was just a small seed. Your belief in its value, your sharp and constructive feedback, and your patience and presence throughout this journey have been instrumental in shaping this thesis. Your thoughtful supervision allowed me to grow in confidence as a researcher and stay connected to the purpose behind the work.

To my family, your support has been unwavering. My Dad and brothers regular check-ins, and Mum, your tireless help in promoting the study to community groups and your constant reminders to care for myself have kept me going. You've both raised me to break barriers, and

this work is a reflection of that. And my close friends, thank you for the memes, the laughter, the timely distractions, and the encouragement. Your love and solidarity have meant the world.

And to my incredible fiancé, Ben, thank you for being my rock through it all. Your calm presence, insightful feedback, and unwavering belief in me have carried me through the most challenging moments. The incredible meals you've prepared and the many thoughtful ways you supported me behind the scenes did not go unnoticed. I am deeply grateful for your continuous support and encouragement.

Lastly, I would like to acknowledge all those who are working to improve equity and inclusion in health care and research. Their efforts have helped pave the way for projects like this one, and I hope this thesis contributes meaningfully to that broader conversation.

Part 1: Literature Review

A Systematic Review of cognitive tools used to assess the prevalence of Dementia in Sub-Saharan Africa with those of African Origin.

Abstract

Aim:

This systematic review aimed to evaluate the cognitive screening tools used to estimate the prevalence of dementia in Sub-Saharan Africa (SSA), with a focus on their psychometric quality and cultural relevance.

Method:

A comprehensive search was conducted in PsycINFO, Medline, and Web of Science using terms related to dementia, cognitive screening, and African populations. Eligible prevalence studies reporting dementia prevalence rates were identified, and cognitive screening tools used to generate those estimates were extracted for further evaluation. Validation studies for these tools were identified through reference list screening, citation tracking, and additional searches in databases such as Google Scholar and UCL Explore. Tools were classified according to levels of cultural adaptation and appraised using the Terwee et al. (2007) criteria and findings across seven psychometric domains were synthesised narratively.

Results:

Twelve different cognitive screening tools were identified from nine prevalence studies, involving 9,957 participants across seven SSA countries. The Community Screening Interview for Dementia (CSI-D) was the most frequently used, followed by the Brief CSI-D and MMSE. Quality appraisal showed that the Rowland Universal Dementia Assessment Scale (RUDAS) had the strongest psychometric and contextual performance, while the MMSE had the weakest. Five tools had been validated in SSA; the remaining six had not, and validation studies from other settings were used to inform appraisal.

Conclusions:

Findings highlight a need for more culturally appropriate and psychometrically robust tools validated within SSA. Clinical implications include the importance of selecting validated tools such as the RUDAS. In the UK, where populations may also reflect cultural and linguistic diversity, clearer guidance, such as through the NICE dementia guidelines, could support clinicians in selecting appropriate tools. Future research should prioritise African-led tool development, integration of such tools into prevalence studies, and the establishment of local normative data to enhance diagnostic accuracy and equity.

Introduction

Background

Dementia is an umbrella term for a range of progressive neurodegenerative conditions that can cause significant cognitive decline, impairing memory, executive function, language, mood and daily activities. These impairments contribute to substantial behavioural and social challenges, significantly impacting the quality of life for individuals and their caregivers (World Health Organisation, 2023). Currently, dementia is the the leading cause of disability and dependence among older adults globally. With no curative treatment available, early detection and intervention remain crucial for mitigating the disease's impact.

The burden of dementia is disproportionately high in low- and middle-income countries (LMICs), where more than 60% of individuals living with dementia reside (Alzheimer's Disease International, 2015). With rising life expectancies and an unprecedented increase in its older population, Sub-Saharan Africa (SSA) is experiencing a growing prevalence of dementia,

making it a major public health challenge. By 2030, the number of people with dementia in Africa is expected to rise by 88%, placing further pressure on already strained healthcare systems (Alzheimer's Disease International, 2015). Despite this growing public health challenge, dementia remains widely underdiagnosed in the region, largely due to limited awareness, inadequate healthcare infrastructure, and a heavy reliance on informal family caregiving (Akinyemi et al., 2022; Paddick et al., 2013).

Although some African nations have introduced policies to support older adults, dementia care remains underdeveloped due to late diagnoses, inadequate workforce training, and limited healthcare accessibility (Alzheimer's Disease International, 2017). As early diagnosis plays a crucial role in planning interventions, public health policies, and resource allocation, the accuracy and reliability of dementia prevalence estimates are of paramount importance.

A systematic review by Mavrodaris et al (2013) sought to estimate the prevalence of dementia and cognitive impairment across Sub-Saharan Africa (SSA), synthesising findings from 19 studies conducted in five countries involving over 10,500 participants. Reported dementia prevalence ranged from 0% to 10.1%, while cognitive impairment rates varied from 6.3% to 25%. Notably, most included studies were conducted in Nigeria, raising concerns about the generalisability of the findings to the broader SSA region. The review identified older age and female sex as consistent risk factors and highlighted substantial variability in prevalence estimates across different countries, settings, and methodological approaches. While the authors identified the screening tools (Community Screening Interview for Dementia, CSI-D; Mini-Mental State Examination, MMSE; 10-Word Delay Recall Test, 10-WDRT; Short Cognitive Evaluation Battery, SCEB; Self-Reporting Questionnaire, SRQ-24; Consortium to Establish for Alzheimer Disease, CERAD; Blessed Dementia Rating Scale, BDS) and diagnostic criteria

(Diagnostic and Statistical Manual of Mental Disorders Fourth Edition, DSM-IV; International Classification of Diseases Tenth Revision, ICD-10; National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association II, NINCDS-ADRA II; and structured clinical assessment) used to assess dementia and cognitive impairment, they did not systematically examine the psychometric properties or cultural relevance of the cognitive tools employed. This methodological omission limits the interpretability and comparability of prevalence estimates and underscores the need for a more critical evaluation of the tools used in dementia research within SSA. Importantly, the authors acknowledged the need for culturally adaptable screening instruments and called for further research to address significant gaps in the existing literature, particularly in southern and eastern Africa, where data remains sparse.

The Role of Cognitive Screening in Dementia Prevalence Studies

Dementia prevalence is primarily estimated through cross-sectional studies, where cognitive screening tools are widely used to identify individuals living with cognitive impairment in large populations (Cleret de Langavant et al., 2020). These tools are essential in epidemiological research, as they allow for efficient, cost-effective, and large-scale dementia assessments. However, the effectiveness of these tools depends on their validity and reliability for the target population (Kamalzadeh et al., 2024).

Many cognitive screening tools used in SSA have been developed in Western contexts and subsequently applied to African populations without adequate cultural and linguistic adaptation. This raises concerns about their applicability, as differences in literacy levels, linguistic diversity, and cultural norms may impact test performance and lead to measurement bias (Cleret de Langavant et al., 2020). A lack of appropriate cognitive screening tools can

result in inaccurate prevalence estimates, which in turn affect policy planning, resource allocation, and the development of dementia care services.

Cognitive Screening Tools in SSA

The lack of consensus on the most appropriate cognitive screening tools for SSA populations has led to substantial variability in dementia prevalence estimates (Alzheimer's Disease International, 2017). Although not formally standardised, many studies discussing cultural considerations in cognitive tests such as Fernandez and Abe (2018), refer to tools as *culturally adapted, newly developed* or *assumed culture-fair* (Chithiramohan et al., 2024; Fernandez & Abe, 2018).

Culturally adapted tools are standard cognitive assessments modified to enhance their linguistic, educational and cultural relevance for a particular population. This can involve replacing culturally unfamiliar test items and translation (Fernandez & Abe, 2018). For example, the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) was adapted for low literacy settings in Tanzania and Nigeria, demonstrating strong reliability and diagnostic accuracy (Paddick et al., 2017), and widely used tools such as the Mini-Mental State Examination (MMSE) have been translated in different African languages (Paddick et al., 2021). In this context, these tools attempt to bridge the gap between Western-developed assessments and African populations by incorporating local norms and cultural considerations.

Newly developed tools are instruments specifically designed from the ground up to reflect their linguistic, cultural, and educational realities of the populations they intended to serve. (Akinyemi et al., 2016; Fernandez & Abe, 2018). In SSA, the Identification of Dementia in Elderly Africans (IDEA) screen was developed in Tanzania and validated in Nigeria, offering a brief and culturally appropriate alternative to more biased measures (Paddick et al., 2015).

The Test of Senegal is another example of an SSA-specific screening tool, incorporating local concepts and languages (Toure et al., 2008). Such tools provide a more ecologically valid method of cognitive assessment**.**

Assumed culture-fair tools are designed to minimise the influence of culture, language, or education, often relying on nonverbal tasks or universally applicable concepts. (Ardila, 2005; van de Vijver & Tanzer, 2004). The Rowland Universal Dementia Assessment Scale (RUDAS) is one which was developed in Australia for multicultural populations and validated across several continents, including Africa (Rowland et al., 2004). While these tools are designed to be widely applicable, they still require local validation to ensure they perform reliably in SSA contexts.

Aim and rationale

While systematic reviews have examined the diagnostic accuracy of cognitive screening tools in other regions, such as Iran (Kamalzadeh et al., 2024) or have broadly assessed dementia screening in illiterate and low-educated populations in LMICs (Paddick et al., 2017), there remains a gap in the literature specific to SSA. Given these challenges, the aim of this systematic review is to identify which cognitive screening tools are currently used in SSA and evaluate their validity, diagnostic accuracy, and cultural appropriateness for dementia screening in these populations.

Methods

Search Strategy development

A rapid scoping review was conducted using Google Scholar, UCL Explore, and PROSPERO to identify existing systematic reviews on cognitive assessments for dementia in minoritised

ethnic communities. This preliminary search provided insights into gaps in the literature and helped refine the search terms and conceptual groupings for the systematic review (e.g., dementia, assessing cognition, geographical terms for African countries).

Search terms

Search terms were developed based on five key concepts: (1) dementia, (2) cognitive assessment, (3) general neuropsychological tests, (4) specific screening tools, and (5) the African population. For the dementia concept, a sample search string included: "dement* OR Alzheimer* OR fronto temporal dement* OR frontotemporal dement* OR vascular dement* OR lewy bod*".

To identify studies involving cognitive assessment, search terms included general phrases such as "assess OR tool OR cognitive screen* OR cognitive assessment OR memory test OR neuropsychological assessment"*. In addition, a separate concept was constructed for general neuropsychological test batteries, using database-specific indexing terms such as "exp neuropsychological tests/".

Named cognitive screening tools were captured using terms like "ACE OR Addenbrooke OR MoCA OR Montreal Cognitive Assessment OR MMSE OR Mini-Mental State Examination OR RUDAS OR Rowland Universal Dementia Assessment Scale"*. In this case, OR was used exclusively to avoid excluding relevant tools not individually named in the search.

To define the target population, a comprehensive list of all 54 African countries was included as individual search terms (e.g., "Nigeria", "Kenya", "Senegal", etc.), alongside broader population descriptors such as "Africa", "ethnic*", "ethnic minority", "BAME", and "BME"*. These terms were combined to form the final concept of the African population. Although the

review focused on Sub-Saharan Africa, all African countries were included in the initial search to maximise sensitivity and account for inconsistent geographic indexing in databases.

Consultation with an academic librarian guided the search approach. The searches were limited to title, abstract, subject headings, tests and measures, and key concepts. This approach was done to enhance the precision of the search results in order to maximise relevance and reduce retrieval of unrelated studies, while still capturing key terms likely to appear in prevalence studies. There were no restrictions on publication year. To ensure comprehensive coverage, citation tracking and reference list screening of included studies and relevant systematic reviews were conducted. Grey literature and preprints were not included due to the focus on peer-reviewed studies with complete methodological reporting.

The final search strategy involved combining all concepts using Boolean logic, with full search strings available in Appendix A.

Databases

The search strategy was applied to Medline, PsycINFO, and Web of Science. These databases were selected due to their extensive coverage of medical, psychological, and interdisciplinary research relevant to cognitive screening and dementia prevalence. Additionally, they were frequently used in previous systematic reviews on similar topics, ensuring methodological alignment with established literature.

Screening and study selection

The study selection process followed the PRISMA 2020 guidelines (Page et al., 2021). The initial search results were imported into Rayyan software (Ouzzani et al., 2016), where duplicate studies were first identified and removed.

Studies were included if they met the following criteria:

- Reported on the prevalence of dementia in a specific country or region within sub-Saharan Africa (SSA).
- Involved adult participants aged 18 years and older of African origin. Where not explicitly stated, studies conducted in SSA were assumed to involve participants of African origin unless otherwise specified.
- Provided quantitative data derived from neurocognitive screening and assessment tools.

Studies were excluded if they met any of the following conditions:

- Were systematic reviews, narrative reviews, or other forms of secondary analysis.
- Did not clearly specify the participant population.
- Used tools that did not assess neurocognitive function (e.g., tools assessing only emotional, behavioural, or physical domains).
- Focused on neurocognitive screening unrelated to dementia.
- Were not published in English, unless an official translation was available.
- Lacked availability of the title, abstract, or full text.
- Were conducted in North African countries (e.g., Egypt, Tunisia, Algeria), due to their distinct sociocultural, linguistic, and historical context, which aligns more closely with the Middle East and North Africa (MENA) region than with SSA.

The total records were retrieved from the three databases PsycINFO, Medline and Web of Science and duplicate records were removed. Thesis author (TO) screened the titles and abstracts for eligibility, and a second reviewer (MG) independently reviewed 10% of the total records. Any discrepancies between MG and TO's screening decisions were resolved through discussion and re-evaluation of the inclusion and exclusion criteria.

Data extraction

Following full-text screening, data were extracted into a Microsoft Excel spreadsheet to capture key study characteristics relevant to the systematic review. For each included study, information was recorded on study characteristics, participant characteristics, and the dementia screening tools used. Where available, information on cultural considerations made were extracted, as well as the reported dementia prevalence rates.

This information was tabulated to form Table 1, which included the following columns: Author/Year, Country and Setting, Sample Size, Participant characteristics (Age, Sex, Education), Screening phases/ Dementia Screening Tools Used/Cut Off Scores, Reference Test, Cultural Considerations and Dementia Prevalence. This structure enabled a comparative analysis of the dementia screening tools employed across studies, as well as similarities and differences in the contexts in which they were applied.

Data synthesis

A narrative synthesis was conducted to compare and interpret the extracted data systematically.

The synthesis involved a descriptive overview of included studies, summarising the cognitive tools used, the cultural considerations, and the study settings. A comparative analysis was

then conducted to categorise screening tools based on their level of cultural adaptation (culturally adapted, newly developed, or assumed culture-fair) and to evaluate their applicability across different populations.

Screening and diagnostic approaches were examined, distinguishing between single-stage and two-stage methods, with consideration of whether clinical reference standards (e.g., DSM-IV) were used. General trends of the dementia prevalence rates were also synthesised.

Potential methodological challenges and biases were considered, including the impact of education, literacy, and linguistic adaptation on screening outcomes. The findings were interpreted to assess how cognitive screening tools have been applied in SSA and their implications for diagnostic accuracy.

Quality assessment of dementia screening tools

To select an appropriate quality assessment tool, the comparative analysis by Rosenkoetter and Tate (2017) was reviewed, which evaluates various critical appraisal tools against the Consensus-based Standards for the selection of health Measurement Instrument (COSMIN) checklist. An initial scoping review was conducted to examine the quality assessment tools used in similar systematic reviews. Based on this review, the Terwee et al. (2007) quality criteria tool was selected due to its comprehensive evaluation of the psychometric properties of health measurement instruments.

The Terwee et al. (2007) tool assesses eight key measurement properties: content validity, internal consistency, criterion validity, construct validity, reproducibility (agreement and reliability), responsiveness, floor and ceiling effects, and interpretability. Sensitivity refers to a tool's ability to correctly identify individuals with dementia (true positives), while specificity

refers to correctly identifying those without dementia (true negatives) (Boehme et al., 2021; Tsoi et al., 2015). These metrics are considered central to criterion validity within the Terwee framework, as they reflect a tool's diagnostic accuracy. While responsiveness is a key domain in the Terwee et al. (2007) criteria, it was not rated in this analysis, as the primary focus was on identifying tools suitable for use in epidemiological surveys of dementia prevalence, where the aim is to discriminate between individuals with and without dementia at a single point in time. Each domain was evaluated using Terwee et al (2007) specified rating system: positive (+), indeterminate (?), negative (-) or not reported (0). As per Terwee's framework, sensitivity, specificity, and diagnostic accuracy (e.g., AUC/ROC) are key components of criterion validity, reflecting how well a tool can distinguish between individuals with and without dementia. These ratings were used to guide a structured comparison of the strengths and limitations of each tool.

The psychometric properties of the cognitive screening tools included in this review were assessed by identifying validation studies through reference list screening, citation tracking, and additional searches in databases such as Google Scholar and UCL Explore. Priority was given to validation studies conducted in SSA to ensure relevance to the target population. However, where no SSA validation studies were available, studies from other regions were included, provided they offered comprehensive or recent psychometric data on the identified screening tools.

Results

PRISMA flow

The search strategy identified records across three electronic databases. After duplicate removal, the remaining records underwent title and abstract screening. Of these, full-text

papers were selected for full-text review. After exclusions based on the eligibility criteria and removal of inaccessible papers, nine studies were retained for final analysis. The PRISMA flowchart (Figure 1) provides a detailed breakdown of the study selection process, including the number of records identified, screened, excluded, and retained for final analysis. No additional prevalence studies were identified through citation tracking or reference list screening.

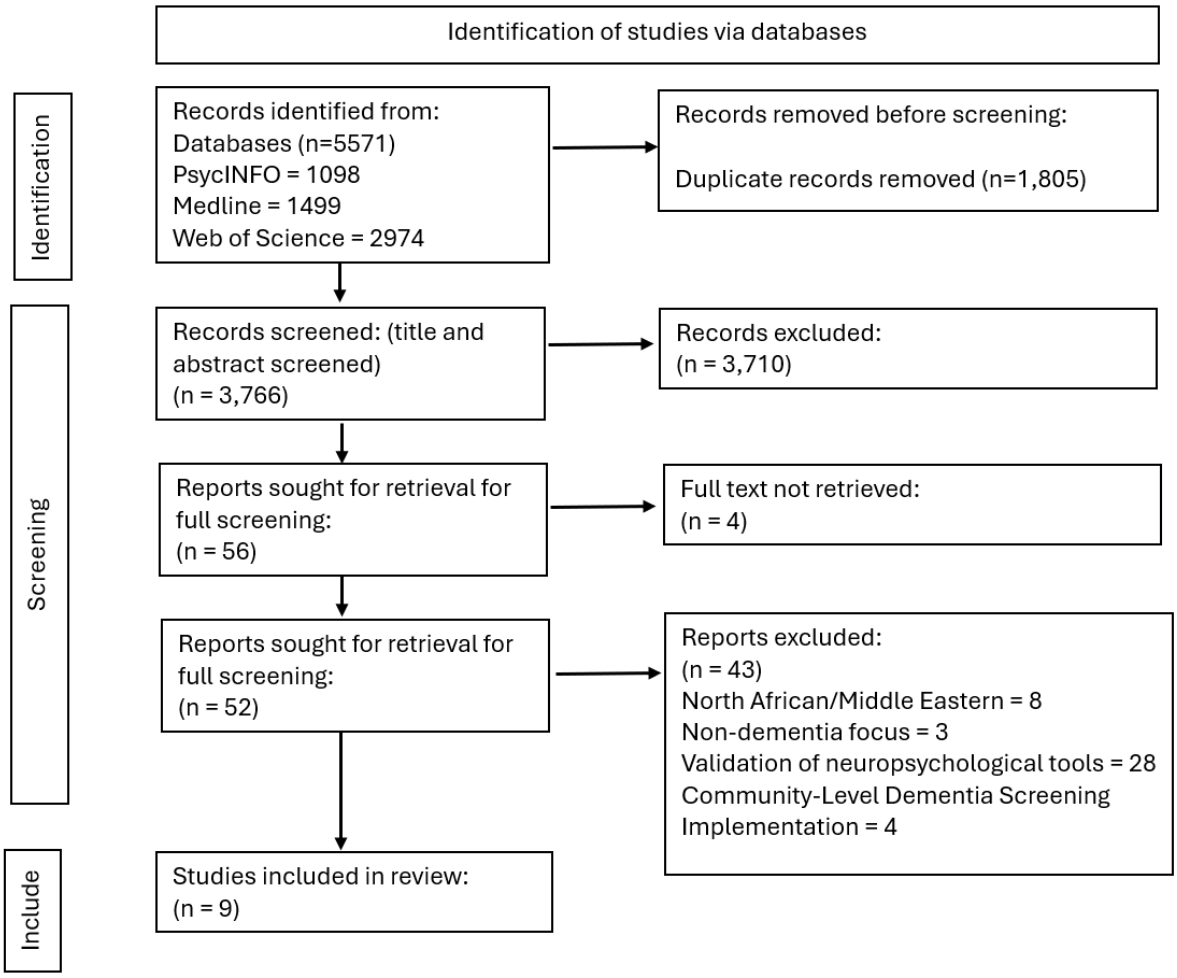


Figure 1.
Prisma flow diagram for identification of SSA dementia prevalence studies.

Table 1. *Summary of key characteristics of prevalence studies, and cultural considerations for cognitive assessment tool.*

1st Key:

Sex: Male (50%) Female (50%),
Screening Phases: single stage ▮, two stage ▮▮
Reference Test: Dementia diagnosis according to DSM-IV criteria ◆, No reference test X
(?): Not stated in paper

Author, Year	Country & Setting	N	Participant characteristics (Age, Sex, Education)	Phases, Tools used and cut off score	Reference test	Cultural considerations	Dementia Prevalence (%)
Adoukonou et al., 2020	Parakou, Benin (urban) Community setting and institutional centres (university hospital)	440	Age: ⩾51 (64.9)  F:7.7%, M: 92.3% Education: 19.6% no formal education, 49.5% primary school level, 30.9% secondary school or higher	▮▮ Brief CSI-D (<7) Below cut off: MMSE (?) CDT (?)	◆ ▮▮ + HACHINSKI ischemic score	**Culturally adapted** *Brief CSI-D translated & back-translated with neurologist review; French version of MMSE adapted by GRÉCO Work Group; all assessments in French; brief CSI-D not validated in Sub-Saharan Africa, acknowledged as a limitation.*	3.2% (95% CI = 1.5 - 4.8)
Awuol, Innocent & Winfred. 2023	Mukono, Uganda (rural) Faith-based geriatric centre	271	Age: ⩾50 (72.2)  F:70.5%, M: 29.5% Education: 50% primary school level	▮ EDQ (⩾8)	X	**Assumed culture fair** *No formal adaptation; Malaysian EDQ used without validation in Uganda; assumed to be culture- & education-independent; study acknowledged this as a limitation.*	46% (95% CI = 0.46; 0.40 - 0.52)
Duodu et al., 2024	Ashanti region of Ghana (urban) Mix of health care facilities	800	Age: ⩾45 (mean age unknown)  F:63.6%, M:36.4% Education: 40.88% secondary school level	▮ RUDAS (<23)	X	**Assumed culture fair** *RUDAS selected as a culturally neutral tool; widely used globally; not validated in Ghana but deemed acceptable for local use*	23.38% (95% CI = 20.44 - 26.31)
Guerchet et al., 2009	Benin, Djija centre (rural) Community setting	502	Age: ⩾65 (76.1)  F:69.1%, M:30.9%	▮▮ CSI – D (<25.5) FWT (<10)	◆ ▮▮	**Culturally adapted** *Trained Fon translators used; CSI-D modified (scoring adjusted, temporal orientation questions*	2.6% (95% CI = 1.1 -3.8)

			Education: 3.4% primary school level, 96.6% illiterate	Below cut off: Oral FCSRT (?) ZCT (?) ISTVF (?)		*removed); clock-drawing test excluded due to high illiteracy; memory tests adapted for oral culture; local normative data developed.*	
Guerchet et al., 2010	Central African Republic (Bangui) and Republic of Congo (Brazzaville) (urban) Community setting	1016	Age (Bangui): ⩾65 (72.7) Age (Brazzaville): ⩾65 (74.7) (Bangui): F:56.1%, M:43.9% (Brazzaville): F:61.9 %, M:38.1 % Education (Bangui): 43.3% primary school + level Education (Brazzaville): 51.1% primary or higher	CSI-D (<23.5/30) FWT (<10/10) Below cut off: FCSRT (?) ZCT (?) ISTVF (?)		**Culturally adapted** *CSI-D & FWT adapted & pretested in Sango, Lingala, Kituba; FWT modified for oral use in illiterate participants; assessments adapted for socio-economic & linguistic diversity; validated historical event-based age estimation used.*	8.1% (95% CI=5.8 - 10.8) 6.7% (95% CI= 4.7 - 9.2) rates close to those observed in high-income countries
Gureje, Ogunniyi & Kola, 2006	Yoruba-speaking states in 8 south western and north-western regions (mixed rural & urban) Community setting	1904	Age: ⩾65 (mean age unknown) Sex: Not stated Education: Not stated	10 – WDRT (⩾2)	X	**Culturally adapted** *Administered in Yoruba Modified CERAD word list (4 words replaced for cultural relevance); demonstrated good cultural appropriateness with minimal education bias; validated with 76.9% sensitivity and 73.5% specificity*	10.1% (95% CI = 8.6-11.8)
Longdon et al., 2013	Hai district of Tanzania (rural) Community setting	1198	Age: ⩾70 (mean age unknown) F:56.2%, M:43.8% Education: 11.8% more than 4years of education	CSI-D (?) Below cut off: CERAD 10-WT (?) GMSE (?)		**Culturally adapted** *All measures translated & back-translated into Swahili; local research nurse assisted interviews; trained local village enumerators; workshops conducted with culturally relevant case histories.*	6.4% (95% CI = 4.9 - 7.9)
Musyimi et al., 2024	Kenya, Makueni county (rural) Community setting	3546	Age: ⩾60 (70.5) F:58.3%, M:41.7%	Brief CSI-D (<5)	Back estimation based on CSI-D established sensitivity and specificity	**Culturally adapted** *All measures forward- & back-translated into Kamba by multilingual experts; reviewed by clinicians, psychologists &*	18.49% (95% CI = 17.1 - 19.7) After back-estimated prevalence.

			Education: 67.9% able to read and write			*community health officers for meaning preservation.*	9.4% (95% CI= 7.9 - 1.8)
Ochayi & Thacher. 2006	Central Nigeria (mix rural & urban) Community setting	280	Age: ⩾65 (73)  F:51.4%, M:48.6% Education: 24% primary school level	CSI-D (⩽28.5)	X	**Assumed culture fair** *CSI-D designed to be culturally & education-independent; previously validated in Nigeria & other developing countries; cut-off score (⩽28.5) validated for Nigerian population.*	6.4% (95% CI = 3.8-9.9)

2nd key: CSI – D, Community Screening Interview for Dementia; Brief CSI-D, Brief Community Screening Interview for Dementia; FWT, Five-Words Test; ZCT, Zazzo cancellation task; ISTVF, Isaac's Set Test of Verbal Fluency; RUDAS, Rowland Universal Dementia Assessment Scale; oral FCSRT, oral Free and Cued Selective Reminder Test; 10 WDRT, 10-Word Delay Recall Test; EDQ, Early Dementia Questionnaire; MMSE, Mini-Mental State Examination; CDT, Clock Drawing Test; CERAD 10-WT, Consortium Establish a Registry for Alzheimer's Disease 10- word task; GMSE, Geriatric Mental State Examination.

Descriptive Overview of Study characteristics

The nine included studies were conducted across seven SSA countries, with a combined total of 9,957 participants. All studies employed a cross-sectional design and were primarily community-based. Sample sizes ranged from N=271 to N=3,546 capturing both small-scale and large population-based cohorts. Most studies included a greater proportion of women, consistent with demographic trends in ageing populations. Participants were generally 60 years and older, though inclusion criteria ranged from 45+ to 70+ across studies. Where reported, education levels were typically low, especially in rural areas, and this context likely shaped cognitive performance and tool selection across the studies.

Across the nine included prevalence studies, twelve different dementia screening tools were identified. The most frequently used was the Community Screening Interview for Dementia (CSI-D), appearing in three studies, followed by the Brief CSI-D, Zazzo Cancellation Task (ZCT), Five-Word Test (FWT), and Isaac's Set of Verbal Fluency (ISTVF), each used in two studies. The remaining tools, including the Early Dementia Questionnaire (EDQ), Rowland Universal Dementia Assessment Scale (RUDAS), Geriatric Mental State Examination (GMSE), Mini-Mental State Examination (MMSE), 10-Word Delayed Recall Test (10-WDRT), CERAD 10-Word Test (CERAD 10-WT), and Clock Drawing Test (CDT)- were each used in one study. This distribution reflects the dominant role of the CSI-D tool used in SSA, while also highlighting the methodological diversity and lack of standardisation across studies.

Most studies were conducted in community-based settings, such as rural or urban households (e.g., Guerchet et al., 2009; Musyimi et al., 2021), except for Awuol et al (2023), which recruited from a faith-based geriatric centre in Uganda and Duodu et al (2020) using a sample from a mix of healthcare facilities in Ghana, this variability in setting likely influenced sample

composition and dementia prevalence rates. In terms of diagnostic approach, Guerchet et al (2009, 2010), Adoukonou et al., 2020 and Longdon et al (2013) used two-stage procedures, combining dementia screening with clinical diagnostic criteria according to the DSM-IV. The remaining applied single-stage screening, classifying dementia based on cut-off scores from tools like the Brief CSI-D and MMSE without further neuropsychological assessment or diagnostic criteria.

Cultural Framing and Adaptation of Cognitive Tools

Based on Fernandez & Abe (2018) explanation of culturally adapted, newly developed and culture-fair tools, five tools were categorised as culturally adapted (e.g., CSI-D, Brief CSI-D, GMSE, FWT, 10-WDRT), though in most cases adaptation was limited to translation. One tool (RUDAS) was explicitly designed to reduce cultural and educational bias and was classified as culture-fair. Three others (EDQ, CDT, ISTVF) were assumed to be culture-fair due to their low language or nonverbal format, though there was no clear evidence of a culture-minimising design. None of the included tools were newly developed for SSA populations. The CSI-D was considered culture-fair in one study conducted in Nigeria, where it was originally developed, for use in Yourba-speaking populations in Nigeria (Hall et al.,1993). However, in other contexts where it was translated (e.g., Kenya, Tanzania, Benin) it was categorised as culturally adapted. This highlights how classification can vary depending on context and use.

Reports of adaptation processes were inconsistent across studies. Guerchet et al (2009, 2010) provided detailed accounts of translating, simplifying and modifying the CSI-D and other memory tasks to suit local literacy levels and sociocultural contexts in Benin and the Central African Republic, including removing culturally unfamiliar items. Musyimi et al (2021) also described systematic translation and back-translation procedures for the Brief CSI-D in Kenya,

alongside pilot testing and community engagement to ensure acceptability. Similarly, Longdon et al (2013) adapted the CSI-D and CERAD 10-WT for use in Tanzania, including translation into Swahili and using culturally appropriate cues. In contrast, Awuol et al (2023), using the EDQ and Duodu et al (2020), using the RUDAS, did not report any adaptation procedures, despite applying dementia screening tools developed in non-African contexts, possibly due to the assumptions of it being culture-fair. These inconsistencies limit the ability to assess the cultural suitability of some tools and raise concerns about bias across diverse study populations.

Variation in Prevalence Rates

Reported dementia prevalence rates varied widely from 2.6% in Benin (Guerchet et al., 2009) to 46% in Uganda (Awuol et al., 2023). Studies using the CSI-D, even when applied in similar community settings, reported varying prevalence rates, for instance, Guerchet et al. (2009) in Benin found 2.6%, while Longdon et al. (2013) in Tanzania reported 6.4%. Such variation may reflect actual differences in population health and demographics but may also be due to methodological differences such as different cut-off scores in the CSI-D (e.g. Guerchet et al 2009 had a cut-off score of <25.5 compared to Guerchet et al 2010, which was <23.5 and to Longdon et al 2013 which did not report the cut-off score). Studies that used single-stage tools without clinical confirmation (e.g., Duodu et al., 2020; Awuol et al., 2023) tended to report higher prevalence rates, raising concerns about potential overestimation without reference standards.

Table 2. *Quality Assessment of Dementia Screening Tools.*

Dementia screening tools & Citation	Content Validity	Internal Consistency	Criterion Validity	Construct Validity	Reproducibility		Floor to Ceiling Effect	Interpretability
					Agreement	Reliability		
RUDAS (Daniel et al, 2020 - Ethiopia)	+	+	+	+	+	0	0	+
ISTVF (Isaacs & Kennie, 1973 - Scotland)	+	0	+	?	0	0	+	?
Oral FCSRT (Montesinos et al, 2022 – Peru)	+	+	+	0	0	+	0	?
EDQ (Arabi et al, 2016 - Malaysia)	+	+	+	?	0	+	0	0
CDT (Emek-Savasm Yerlikaya & Yener, 2018 – Turkey)	+	-	+	0	+	+	?	+
GMSE (Prince et al, 2004 - Nigeria)	+	+	+	0	0	0	0	?
CSI – D (Hall et al, 2000 - Nigeria)	+	0	+	0	0	+	0	?
FWT (Mormount et al, 2012 – Belgium)	+	0	+	+	-	-	?	+
CERAD 10 WT (Unverzagt et al., 1999 - Jamaica)	+	0	+	+	0	0	-	+
Brief CSI – D (Prince et al, 2011 – Nigeria)	+	?	+	0	0	0	0	-
MMSE (Vissoci et al, 2019 – Tanzania)	+	-	0	-	0	0	0	0

Key: + = positive rating; ? = indeterminate rating; - = negative rating; 0 = no information available

RUDAS, Rowland Universal Dementia Assessment Scale; ISTVF, Isaac's Set Test of Verbal Fluency; Oral Free and Cued Selective Reminder Test; EDQ, Early Dementia Questionnaire; CDT, Clock Drawing Test; GMSE, Geriatric Mental State Examination; CSI – D, Community Screening Interview for Dementia; FWT, Five-Words Test; CERAD 10-WT, Consortium Establish a Registry for Alzheimer's Disease 10- word task; Brief CSI-D, Brief Community Screening Interview for Dementia; MMSE, Mini-Mental State Examination.

Summary of Tools Included in the Quality Assessment

The quality assessment of dementia screening tools identified from the included prevalence studies was conducted using the Terwee et al (2007) criteria. To ensure contextual relevance, priority was given to validation studies conducted in SSA. Studies from other regions were included where SSA validation studies were unavailable, provided they offered comprehensive or recent psychometric evaluation. Of the tools identified, two were excluded from the quality assessment due to lack of accessible validation studies: the 10-Word Delayed Recall Test (10-WDRT) had no identifiable validation papers, and the Zazzo Cancellation Task (ZCT) was omitted due to its full text being unavailable in English.

A total of 11 tools were included in the quality assessment. Five tools were validated in African countries: the RUDAS in Ethiopia, the GMSE, Brief CSI-D and CSI-D, in Nigeria, and the MMSE in Tanzania. The remaining six tools were included based on validation studies conducted outside Africa: the EDQ (Malaysia), ISTVF (Scotland), Oral FCSRT (Peru), CDT (Turkey) FWT (Belgium), and CERAD 10-WT (Jamaica). This mixed inclusion reflects the limited availability of African validation studies for some tools and highlights the need for caution when generalising findings across cultural contexts. For some tools, such as the 10-WDRT and CSI-D, diagnostic accuracy metrics (e.g., sensitivity and specificity) were reported in individual studies (e.g., Gureje et al., 2006; Ochayi & Thacher, 2006; Musyimi et al., 2024), supporting their criterion validity. However, these were not consistently available across tools and did not always align with broader psychometric evaluations. Table 2 presents the quality assessment results.

Performance of Tools Validated in SSA

The quality assessment revealed substantial variation in the psychometric strength and contextual relevance of the dementia screening tools included. The RUDAS showed the

strongest psychometric properties, demonstrating strong evidence with positive ratings across multiple domains including content validity, internal consistency, criterion and construct validity and test-retest reliability. As a tool explicitly designed to be culture-fair, its performance underscores the value of culturally neutral test design when supported by local validation. In contrast, the MMSE showed the weakest psychometric evidence among the cognitive screening tools , with negative ratings for internal consistency and construct validity, alongside limited supporting data across the remaining domains. Most tools showed gaps in reproducibility and interpretability, particularly in reporting how scores vary across different subgroups (e.g. by age, education or gender) and in defining score thresholds that indicate clinically meaningful impairment.

Among the five tools validated in African populations, only the RUDAS, tested in Ethiopia, demonstrated both high methodological quality and contextual applicability. It was the only tool to achieve positive ratings across nearly all psychometric domains assessed, reflecting a strong evidence base for its reliability and validity in SSA contexts. The GMSE, CSI-D, and Brief CSI-D, although widely used and validated in studies that included Nigerian populations, were supported by fewer psychometric properties and lacked formal evidence for reliability or interpretability, limiting the confidence with which results from these tools can be applied to diverse subgroups or used in tracking population-level patterns. These culturally adapted tools showed moderate psychometric strength but were limited by incomplete validation, especially regarding reproducibility and responsiveness. The MMSE, despite its Tanzanian validation, showed weak performance overall, receiving negative ratings on key psychometric domains. This underscores the limitations of relying on tools that were originally developed in Western contexts, known to be influenced by literacy, education and cultural familiarity. Taken together, these findings suggest that although several tools have been used or culturally

adapted in African studies, their methodological issues constrain their ability for these tools to produce accurate and culturally appropriate estimates of dementia prevalence in SSA. Overall, there remains a lack of screening tools that are both rigorously validated and relevant to the local context, highlighting the need for further high-quality, culturally informed psychometric research in the region.

Performance of Tools Validated outside SSA

The six tools not validated in Africa showed mixed psychometric properties. The EDQ, Oral FCSRT and CDT showed relatively strong validity in their respective studies in Malaysia, Peru and Turkey but their lack of African validation raises issues on its cross-cultural applicability in SSA. The CDT in particular, showed strengths in content and criterion validity, as well as reliability and agreement, but lacked construct validity and showed limitations in internal consistency. Similarly, the CERAD 10-WT and FWT and ISTVF, although validated in Jamaica, Belgium, and Scotland respectively, showed strengths in specific areas such as construct validity or internal consistency, indicating that the items were internally aligned and conceptually relevant to cognitive functioning. However, the absence of reproducibility data raises concerns about the consistency of these tools across settings which is crucial for reliable prevalence estimates. Among the tools assumed to be culture-fair, psychometric performance varied widely, and those lacking SSA validation offered limited assurance of contextual.

Cultural Suitability and Gaps in Validation

Although several tools were described as "culturally adapted" or "assumed culture-fair," few were developed *with* African populations. In some cases, adaptations involved translation and minor content modification to account for language or literacy, as seen with the CSI-D and GMSE. However, these adaptations were rarely evaluated against the full spectrum of

psychometric standards. Most tools developed outside of SSA lacked evidence of cultural equivalence in terms of item familiarity, socio-linguistic appropriateness, or normative benchmarks.

Overall, the findings highlight a limited pool of dementia screening tools that meet both psychometric and contextual standards for use in SSA. Only the RUDAS (culture-fair), validated within the region, approached this threshold. The remaining tools (culturally-adapted), whether validated within or outside Africa, demonstrate varying levels of methodological quality but often fall short in terms of cultural specificity or comprehensive validation. This underscores the need for further context-sensitive psychometric research to inform dementia screening across diverse African settings.

Discussion

This systematic review evaluated cognitive screening tools used in dementia prevalence studies across SSA, focusing on their methodological quality and cultural relevance. The findings highlight key challenges, including limited cultural adaptation, inconsistent validation, and varied diagnostic protocols, all impacting the accuracy and equity of dementia prevalence estimates across the region.

Interpretation of Findings

This review identified 12 distinct cognitive screening tools used across nine dementia prevalence studies, involving 9,957 participants from seven countries within SSA. The CSI-D emerged as the most frequently utilised screening tool, featuring in three studies, followed by the Brief CSI-D and the MMSE. Psychometric appraisal using the Terwee et al (2007) criteria identified the RUDAS as having the strongest overall performance based on available evidence,

while the MMSE scored lowest across multiple domains. Of the 12 tools, five had undergone validation within SSA, while the remainder relied on data from other regions. These tools included those that were culturally adapted or assumed to be culture-fair but none that were newly developed specifically for SSA populations.

The most frequently used screening tool was the CSI-D and its brief version. These tools were developed through the Indianapolis-Ibadan Dementia Project, a collaboration between Indiana University and the University of Ibadan, to study dementia in African American and Yourba Nigerian populations using culturally sensitive methods (Hall et al., 2000; Hendrie et al.,1995). While widely implemented in SSA, the CSI-D was developed with strong attention to feasibility and cultural appropriateness in cross-national, low-literacy settings (Hall et al., 1993), its validation primarily focused on clinical discrimination such as sensitivity and specificity, rather than comprehensive psychometric evaluation across domains such as reliability, construct validity and responsiveness. This background may help explain the limited validation and psychometric data available for the CSI-D and for many tools of a similar nature when appraised against contemporary standards (Boehme et al., 2021; Terwee et al., 2007; Tsoi et al., 2015).

The present review evaluated a wider range of dementia-specific tools such as the RUDAS, Brief CSI-D, FCSRT, appraising their psychometric properties using the Terwee et al. (2007) criteria. Several of these tools have been validated in SSA or other LMICs. The review also incorporates newer studies published since 2013 and includes research from underrepresented regions such as East and Central Africa. In doing so, it moves beyond prevalence reporting to provide a more critical evaluation of the measurement tools that

underpin such estimates, while also expanding to a wider range of SSA regions, an essential step toward improving diagnostic equity and data reliability in SSA.

These findings build on the earlier work of Mavrodaris et al. (2013), who synthesised dementia prevalence estimates across SSA but did not critically assess the tools used to generate them. Their review identified instruments such as the CSI-D, MMSE, 10-WDRT, and ZCT, along with broader psychiatric tools like the Self-Reporting Questionnaire-24 (SRQ-24) and the Short Cognitive Evaluation Battery (SCEB), which were excluded from the present review due to their limited focus on cognitive screening. Reported prevalence rates ranged from 0% to 10.1% for dementia and 6.3% to 25% for cognitive impairment. In contrast, the current review dementia prevalence rates ranged from 2.6% in Benin (Guerchet et al., 2009) to 46% in Uganda (Awuol et al., 2023) revealing a markedly wider range of prevalence estimates. This divergence may reflect differences in diagnostic methodology, with some studies in the current review relying on single-stage screening without clinical confirmation. Overall, both reviews illustrate considerable variability in prevalence rates across SSA underscoring the continuing challenges in achieving consistent measurement in dementia epidemiology.

Strengths and Limitations of Included Studies

Several included studies demonstrated methodological strengths. Many were conducted in community settings with large and demographically diverse samples, enhancing the generalisability of findings. The studies also spanned seven countries across SSA, incorporating both urban and rural populations and contributing to a broader understanding of dementia epidemiology in the region.

A key limitation across the included studies was the insufficient reporting of psychometric properties for the cognitive screening tools used. The quality assessment using Terwee et al

(2007) revealed widespread gaps, particularly in reproducibility, interpretability, and subgroup performance. In many cases, these gaps reflected either the absence of validation studies in SSA or inadequate reporting standards in available publications. For example, the ZCT and 10-WDRT could not be appraised due to the unavailability of accessible validation studies in English. These may exist in other languages or unpublished formats but were not retrievable through the current review methods. These deficits limited the ability to determine how well tools function in SSA contexts and hinder cross-study comparison. Without transparent and robust psychometric data, it becomes difficult to assess whether observed prevalence rates reflect true population-level differences or measurement artefacts.

Strengths and Limitations of Methodology

A strength of this review was its focus on tools already used in population-level prevalence studies, rather than limiting inclusion to those newly developed or validated within SSA. This "prevalence-first" approach provides a more ecologically valid reflection of real-world dementia screening practices across the region. It also ensures that findings are grounded in tools currently shaping prevalence estimates and clinical decisions in SSA.

However, this approach also introduced limitations. By focusing exclusively on tools used in prevalence studies, the review inadvertently excluded newly developed promising tools that have also been implemented in SSA. For example, the IDEA cognitive screen, designed specifically for low-literacy populations in SSA (Paddick et al., 2015), has been validated in several settings but was omitted because it has not yet been applied in prevalence research and thus did not meet the eligibility criteria for inclusion. This may have led to an underrepresentation of emerging tools with strong contextual relevance but limited diffusion in epidemiological work to date.

While this review used comprehensive search strategies across major academic databases (PsycINFO, Medline, Web of Science), it did not include regional databases, such as 'African Journals Online' or 'Global Health database' due to access limitations. As a result, relevant African-led studies or locally published validation data may have been missed. Furthermore, identification of validation studies relied on targeted citation tracking, Google Scholar, and UCL explore, rather than an additional systematic search across all databases. This may have further contributed to gaps in the psychometric data available for quality appraisal.

Finally, the review revealed substantial variability in how tools were implemented and scored across the included studies. Even when the same tool was used (e.g. CSI-D), differences in cut-off scores or unreported threshold reporting made it difficult to compare prevalence rates reliably. Such variation likely contributed to the wide range of prevalence estimates observed across studies (Paddick et al., 2017; Salmon et al., 1989), and limits the extent to which psychometric quality alone can account for differences in reported outcomes.

Implications

Clinical Practice

Cognitive screening practices in SSA often incorporate informant perspectives, particularly through tools like the CSI-D and GMSE, which use structured interviews with relatives or carers alongside direct cognitive assessment. This reflects a more holistic and socially embedded approach to dementia identification, rooted in the centrality of extended family systems and collective caregiving responsibilities (Duodu et al., 2024). In contrast, dementia assessment in Western contexts often prioritises individualised, performance-based tools that isolate cognitive performance from social context, despite evidence that family input remains critical to early detection (Akinyemi et al., 2022).

This difference in clinical emphasis highlights key factors that can be implemented in both SSA and UK practice. In SSA, further integration of informant-based tools into routine care may align well with cultural values and improve diagnostic relevance. In the UK, for those of African heritage or other cultures that similarly place high value on collective caregiving, there is a case for re-evaluating the dominance of individual cognitive scoring in clinics. Integrating carer perspectives and recognising broader social functioning may support more culturally valid and equitable dementia diagnosis in an increasingly diverse population.

Many prevalence studies sampled populations with low levels of formal education, particularly in rural communities. Education level has a well-documented impact on cognitive test performance (Ganguli et al, 2010), yet few validation studies accounted for this adequately, particularly tools developed in Western contexts that assume familiarity with abstract reasoning or literacy-based tasks. The mismatch between test design, validation sample characteristics, and the populations assessed poses a significant risk of measurement bias and subsequently contributes to potential under-or overdiagnosis.

In light of these findings, clinicians in both SSA and UK memory clinics should prioritise the use of tools that minimise cultural and educational bias and have demonstrated strong psychometric performance in diverse populations. The RUDAS, which received the highest quality appraisal in this review based on the Terwee et al (2007) criteria, represents a particularly suitable option given its emphasis on cross-cultural applicability and reduced reliance on literacy-dependent tasks. Conversely, the MMSE, which showed the weakest psychometric performance, should be used with caution in African populations, where its standardised format may not account for linguistic, educational or cultural difference that influence test performance and risk misclassification.

Research

There is a pressing need to integrate African-developed cognitive screening tools into large-scale research, including prevalence studies, comparisons with more commonly used instruments and cross-cultural validation efforts. Several tools designed specifically for SSA contexts, such as those that did not reach the eligibility criteria in this study such as the IDEA remain underrepresented in epidemiological work. Its inclusion in population-based studies would help build a more robust and contextually grounded evidence base, ensuring that tools tailored for local populations are not marginalised in dementia care policy and planning.

Relatedly, although responsiveness, the ability of a tool to detect meaningful change over time, is a key psychometric property, it was not rated in this review due to its limited relevance to the aims of cross-sectional prevalence studies. However, responsiveness remains an important consideration in other contexts, such as clinical services where tools may be used longitudinally to monitor cognitive decline or track disease progression. It is also essential for intervention studies or trials aiming to evaluate treatment effects. Future psychometric evaluations of cognitive screening tools used in SSA should consider including responsiveness data, particularly for tools intended for repeated administration or integration into diagnostic pathway

Although some tools showed strong psychometric performance, scoring thresholds varied widely across studies and were often based on criteria developed outside of SSA. This inconsistency highlights the need to establish locally derived norms and context-specific thresholds that better reflect the region's educational, linguistic, and cultural diversity. Implementing such standards would help minimise the risk of misclassification, reduce the

risk of over- or underdiagnosis, and improve the comparability of prevalence estimates across settings.

Policy

This review demonstrates that most dementia prevalence studies in SSA have relied on community-based screening approaches, often implemented in rural areas and targeting populations with low levels of formal education. This highlights the value and feasibility of community-led screening in reaching underserved groups. National dementia strategies across SSA should therefore continue to support and strategically invest in community-based screening infrastructure. To build on current practice, policy efforts could focus on promoting the use of tools with stronger psychometric support (e.g., RUDAS) and encourage consistent scoring and diagnostic protocols across regions. Additionally, policies should emphasise the importance of culturally appropriate tools that are validated for SSA populations, particularly those with limited literacy, to ensure fair and accurate diagnoses in public health settings.

In the UK, current NICE guidelines (NICE, 2018), advises that dementia assessments should be appropriate for diverse populations, yet they stop short of endorsing specific tools. This lack of clarity can result in inconsistent practice and perpetuate diagnostic disparities among minoritised ethnic groups, particularly those of African heritage. There is a need for more explicit national guidance that endorses the use of culturally fair screening tools such as the RUDAS and supports their routine implementation in memory services. Aligning assessment practices with the cultural and social realities of the UK's diverse population is a necessary step toward achieving greater diagnostic equity.

Conclusion

Findings highlight a need for more culturally appropriate and psychometrically robust tools validated within SSA. Clinical implications include the importance of selecting validated tools such as the RUDAS. In the UK, where populations may also reflect cultural and linguistic diversity, clearer guidance, such as through the NICE dementia guidelines, could support clinicians in selecting appropriate tools. Future research should prioritise African-led tool development, integration of such tools into prevalence studies, and the establishment of local normative data to enhance diagnostic accuracy and equity.

References

Adoukonou, T., Yoro-Zohoun, I., Gnonlonfoun, D. D., Amoussou, P., Takpara, C., Agbetou, M., Guerchet, M., Preux, P.-M., Houinato, D., & Ouendo, E.-M. (2020). Prevalence of dementia among well-educated old-age pensioners in Parakou (Benin) in 2014. *Dementia and Geriatric Cognitive Disorders, 49*(2), 210–218. https://doi.org/10.1159/000508623

Akinyemi, R. O., Baiyewu, O., Dotchin, C., Ighodaro, A., Ogunniyi, A., & Walker, R. W. (2016). Neuropsychological assessment of older Africans: A review of tests adapted and validated in sub-Saharan Africa. *The Lancet Neurology, 15*(6), 567–575. https://doi.org/10.1016/S1474-4422(16)00002-3

Akinyemi, R. O., Yaria, J., Ojagbemi, A., Guerchet, M., Okubadejo, N., Njamnshi, A. K., Sarfo, F. S., Akpalu, A., Ogbole, G., Ayantayo, T., Adokonou, T., Bobrow, K., Mahlanza, L. M., Ranchod, K., & Kalaria, R. N. (2022). Dementia in Africa: Current evidence, knowledge gaps, and future directions. *Alzheimer's & Dementia, 18*(4), 790–809. https://doi.org/10.1002/alz.12432

Alzheimer's Disease International. (2015). *World Alzheimer report 2015: The global impact of dementia – An analysis of prevalence, incidence, cost and trends*. https://www.alzint.org/u/WorldAlzheimerReport2015.pdf

Alzheimer's Disease International. (2017). *Dementia in sub-Saharan Africa*. https://www.alzint.org/resource/dementia-in-sub-saharan-africa/

Arabi, Z., Syed Abdul Rahman, S. A., Hazmi, H., & Hamdin, N. (2016). Reliability and construct validity of the Early Dementia Questionnaire (EDQ). *BMC Geriatrics, 16*, Article 202. https://doi.org/10.1186/s12877-016-0384-1

Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. *Neuropsychology Review, 15*(4), 185–195. https://doi.org/10.1007/s11065-005-9180-y

Awuol, M. D. A., Innocent, B. K., & Winfred, A. (2023). Prevalence, factors associated and knowledge of probable dementia among adults 50 years and over attending a faith-based geriatric center in Uganda. *Scientific Reports, 13*, Article 6737. https://doi.org/10.1038/s41598-023-33948-9

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Boehme, S., Buerger, K., Teipel, S., & Hampel, H. (2021). *The Bamberg Dementia Screening Test (BDST): Diagnostic accuracy and use in a memory clinic setting*. BMC Medicine, 19(1), 1–12. https://doi.org/10.1186/s12916-021-02146-1

Chithiramohan, T., Santhosh, S., Threlfall, G., Hull, L., Mukaetova-Ladinska, E. B., Subramaniam, H., & Beishon, L. (2024). Culture-fair cognitive screening tools for assessment of cognitive impairment: A systematic review. *Journal of Alzheimer's Disease Reports, 8*, 289–306. https://doi.org/10.3233/ADR-230194

Cleret de Langavant, L., Bayen, É., Bachoud-Lévi, A.-C., & Yaffe, K. (2020). Approximating dementia prevalence in population-based surveys of aging worldwide: An

unsupervised machine learning approach. *Alzheimer's & Dementia: Translational Research & Clinical Interventions, 6*(1), e12074. https://doi.org/10.1002/trc2.12074

Daniel, B., Agenagnew, L., Workicho, A., & Abera, M. (2022). Validation of the Rowlands Universal Dementia Assessment Scale (RUDAS) to detect major neurocognitive disorder among elderly people in Ethiopia, 2020. *PLOS ONE, 17*(1), e0262483. https://doi.org/10.1371/journal.pone.0262483

Duodu, P. A., Dey, N. E. Y., Okyere, J., Simkhada, B., Barker, C., Gillibrand, W., & Simkhada, P. (2024). Gendered differences in the prevalence and associated factors of dementia in Ghana: A cross-sectional survey. *BMC Psychiatry, 24*, Article 397. https://doi.org/10.1186/s12888-024-05856-3

Duodu, P. A., Okyere, J., Simkhada, B., Akrong, R., Barker, C., Gillibrand, W., & Simkhada, P. (2024). Exploration of family caregivers' experiences on coping in dementia care in Ghana: A phenomenological study. *BMC Psychology, 12*, Article 361. https://doi.org/10.1186/s40359-024-01862-y

Emek-Savaş, D. D., Yerlikaya, D., & Yener, G. G. (2018). Validity, reliability and Turkish norm values of the Clock Drawing Test for two different scoring systems. *Turkish Journal of Neurology, 24*(3), 143–152. https://doi.org/10.4274/tnd.26504

Fernández, A. L., & Abe, J. A. A. (2018). Bias in cross-cultural neuropsychological testing: Problems and possible solutions. *Culture and Brain, 6*(1), 1–35. https://doi.org/10.1007/s40167-017-0050-2

Ganguli, M., Snitz, B. E., Lee, C.-W., Vanderbilt, J., Saxton, J. A., & Chang, C.-C. H. (2010). Age and education effects and norms on a cognitive test battery from a population-based

cohort: The Monongahela–Youghiogheny Healthy Aging Team (MYHAT). *Aging & Mental Health, 14*(1), 100–107. https://doi.org/10.1080/13607860903071014

Guerchet, M., Houinato, D., Paraiso, M. N., von Ahsen, N., Nubukpo, P., Otto, M., Clément, J.-P., Preux, P.-M., & Dartigues, J.-F. (2009). Cognitive impairment and dementia in elderly people living in rural Benin, West Africa. *Dementia and Geriatric Cognitive Disorders, 27*(1), 34–41. https://doi.org/10.1159/000188661

Guerchet, M., M'belesso, P., Mouanga, A. M., Bandzouzi, B., Tabo, A., Houinato, D. S., Paraïso, M. N., Cowppli-Bony, P., Nubukpo, P., Aboyans, V., Clément, J.-P., Dartigues, J.-F., & Preux, P.-M. (2010). Prevalence of dementia in elderly living in two cities of Central Africa: The EDAC survey. *Dementia and Geriatric Cognitive Disorders, 30*(3), 261–268. https://doi.org/10.1159/000320247

Gureje, O., Ogunniyi, A., & Kola, L. (2006). The profile and impact of probable dementia in a sub-Saharan African community: Results from the Ibadan Study of Aging. *Journal of Psychosomatic Research, 61*(3), 327–333. https://doi.org/10.1016/j.jpsychores.2006.05.053

Hall, K. S., Gao, S., Emsley, C. L., Ogunniyi, A. O., Morgan, O., & Hendrie, H. C. (1999). Community Screening Interview for Dementia (CSI-D): Performance in five disparate study sites. *International Journal of Geriatric Psychiatry, 14*(5), 411–420. https://doi.org/10.1002/(SICI)1099-1166(199905)14:5<411::AID-GPS933>3.0.CO;2-Z

Hall, K. S., Gao, S., Unverzagt, F. W., & Hendrie, H. C. (2000). Low education and childhood rural residence: Risk for Alzheimer's disease in African Americans. *Neurology, 54*(1), 95–99. https://doi.org/10.1212/WNL.54.1.95

Hall, K. S., Hendrie, H. C., Brittain, H. M., Norton, J. A., Rodgers, D. D., Prince, C. S., Pillay, N., & Osuntokun, B. O. (1993). The development of a dementia screening interview in two distinct languages. *International Journal of Methods in Psychiatric Research, 3*, 1–28.

Hendrie, H. C., Hall, K. S., Ogunniyi, A., & Hui, S. L. (1995). The development of a cross-cultural neuropsychological test battery for a study of dementia in Nigeria and the United States. *The International Journal of Methods in Psychiatric Research, 5*(2), 259–269.

Isaacs, B., & Kennie, A. T. (1973). The set test as an aid to the detection of dementia in old people. *The British Journal of Psychiatry, 123*(575), 467–470. https://doi.org/10.1192/bjp.123.4.467

Kamalzadeh, L., Tayyebi, G., Shariati, B., Shati, M., Saeedi, V., & Malakouti, S. K. (2024). Diagnostic accuracy of cognitive screening tools validated for older adults in Iran: A systematic review and meta-analysis. *BMC Geriatrics, 24*, Article 428. https://doi.org/10.1186/s12877-024-04963-w

Longdon, A. R., Paddick, S.-M., Kisoli, A., Dotchin, C., Gray, W. K., Dewhurst, F., Chaote, P., Teodorczuk, A., Dewhurst, M., & Walker, R. (2013). The prevalence of dementia in rural Tanzania: A cross-sectional community-based study. *International Journal of Geriatric Psychiatry, 28*(7), 728–737. https://doi.org/10.1002/gps.3880

Mavrodaris, A., Powell, J., & Thorogood, M. (2013). Prevalences of dementia and cognitive impairment among older people in sub-Saharan Africa: A systematic review. *Bulletin of the World Health Organization, 91*(10), 773–783. https://doi.org/10.2471/BLT.13.118422

Montesinos, R., Parodi, J. F., Diaz, M. M., Herrera-Perez, E., Valeriano-Lorenzo, E., Soto, A., Delgado, C., Slachevsky, A., & Custodio, N. (2022). Validation of picture free and cued selective reminding test for illiteracy in Lima, Peru. *American Journal of Alzheimer's Disease & Other Dementias,* *37*(2), 1–14. https://doi.org/10.1177/15333175221094396

Mormont, E., Jamart, J., & Robaye, L. (2012). Validity of the Five-Word Test for the evaluation of verbal episodic memory and dementia in a memory clinic setting. *Journal of Geriatric Psychiatry and Neurology,* *25*(2), 78–84. https://doi.org/10.1177/0891988712445088

Musyimi, C. W., Ndetei, D. M., Muyela, L. A., Masila, J., & Farina, N. (2024). Dementia screening in rural Kenya: The prevalence and impact of screening positive for dementia. *Neuroepidemiology, 58*(3), 174–181. https://doi.org/10.1159/000536012

NICE. (2018). Recommendations | Dementia: assessment, management and support for people living with dementia and their carers | Guidance | NICE. Nice.org.uk; NICE. https://www.nice.org.uk/guidance/ng97/chapter/Recommendations#diagnosis

Ochayi, B., & Thacher, T. D. (2006). Risk factors for dementia in central Nigeria. *Aging & Mental Health, 10*(6), 616–620. https://doi.org/10.1080/13607860600736182

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews,* *5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Paddick, S. M., Gray, W. K., Ogunniyi, A., Kisoli, A., Dotchin, C. L., & Walker, R. W. (2015). Validation of the Identification and Intervention for Dementia in Elderly Africans (IDEA)

cognitive screen in Nigeria and Tanzania. *International Psychogeriatrics, 27*(4), 717–726. https://doi.org/10.1017/S1041610214002562

Paddick, S.-M., Gray, W. K., McGuire, J., Richardson, J., Dotchin, C., & Walker, R. W. (2017). Cognitive screening tools for identification of dementia in illiterate and low-educated older adults: A systematic review and meta-analysis. *International Psychogeriatrics, 29*(6), 897–929. https://doi.org/10.1017/S1041610216001976

Paddick, S.-M., Kisoli, A., Mkenda, S., Mbowe, G., Gray, W. K., Dotchin, C., Ogunniyi, A., Kisima, J., Olakehinde, O., Mushi, D., & Walker, R. W. (2017). Adaptation and validation of the Alzheimer's Disease Assessment Scale – Cognitive (ADAS-Cog) in a low-literacy setting in sub-Saharan Africa. *Acta Neuropsychiatrica, 29*(4), 244–251. https://doi.org/10.1017/neu.2016.65

Paddick, S.-M., Longdon, A. R., Kisoli, A., Dotchin, C. L., Gray, W. K., Dewhurst, F., Chaote, P., & Walker, R. W. (2015). The identification of dementia in the community in Tanzania: A study comparing the diagnostic accuracy of the IDEA cognitive screen and the CSI-D. *International Journal of Geriatric Psychiatry, 30*(10), 970–977. https://doi.org/10.1002/gps.4245

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development Perspectives, 1*(2), 131–134. https://doi.org/10.1111/j.1750-8606.2007.00020.x

Prince, M., Acosta, D., Chiu, H., Copeland, J., Dewey, M., Scazufca, M., & Varghese, M. (2004). Effects of education and culture on the validity of the Geriatric Mental State and its AGECAT algorithm. *The British Journal of Psychiatry, 185*(5), 429–436. https://doi.org/10.1192/bjp.185.5.429

Prince, M., Acosta, D., Ferri, C. P., Guerra, M., Huang, Y., Jacob, K. S., Llibre Rodriguez, J. J., Salas, A., Sosa, A. L., Williams, J. D., & Hall, K. S. (2011). A brief dementia screener suitable for use by non-specialists in resource-poor settings: The cross-cultural derivation and validation of the brief Community Screening Instrument for Dementia. *International Journal of Geriatric Psychiatry, 26*(9), 899–907. https://doi.org/10.1002/gps.2622

Rosenkoetter, U., & Tate, R. L. (2017). Assessing features of psychometric assessment instruments: A comparison of the COSMIN checklist with other critical appraisal tools. *Brain Impairment, 19*(1), 103–118. https://doi.org/10.1017/BrImp.2017.29

Rowland, J. T., Storey, J. E., Basic, D., Conforti, D. A., & Dickson, H. G. (2004). The Rowland Universal Dementia Assessment Scale (RUDAS): A multicultural cognitive assessment scale. *International Psychogeriatrics, 16*(1), 13–31. https://doi.org/10.1017/S1041610204000043

Salmon, D. P., Riekkinen, P. J., Katzman, R., et al. (1989). Cross-cultural studies of dementia: A comparison of Mini-Mental State Examination performance in Finland and China. *Archives of Neurology, 46*(7), 769–772. https://doi.org/10.1001/archneur.1989.00520430113031Storey, J. E., Rowland, J. T. J., Basic, D., & Conforti, D. A. (2004). The Rowland Universal Dementia Assessment Scale

(RUDAS): A multicultural cognitive assessment scale. *International Psychogeriatrics, 16*(1), 13–31. https://doi.org/10.1017/S1041610204000043

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Touré, K., Coumé, M., Ndiaye-Ndongo, N. D., Sarr, M. M., & Sow, A. D. (2008). The Test of Senegal: A valid and reliable screening tool to assess for dementia in a Senegalese elderly population. *African Journal of Neurological Sciences, 27*(1), 4–13.

Tsoi, K. K. F., Chan, J. Y. C., Hirai, H. W., Wong, S. Y. S., & Kwok, T. C. Y. (2015). *Cognitive tests to detect dementia: A systematic review and meta-analysis*. JAMA Internal Medicine, 175(9), 1450–1458. https://doi.org/10.1001/jamainternmed.2015.2150

Unverzagt, F. W., Morgan, O. S., Thesiger, C. H., Eldemire, D. A., Luseko, J., Pokuri, S., Hui, S. L., Hall, K. S., & Hendrie, H. C. (1999). Clinical utility of CERAD neuropsychological battery in elderly Jamaicans. *Journal of the International Neuropsychological Society, 5*(3), 255–259. https://doi.org/10.1017/S1355617799003082

van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology, 54*(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004

Vissoci, J. R. N., de Oliveira, L. P., Gafaar, T., Haglund, M. M., Mvungi, M., Mmbaga, B. T., & Staton, C. A. (2019). Cross-cultural adaptation and psychometric properties of the MMSE and MoCA questionnaires in Tanzanian Swahili for a traumatic brain injury

population. *BMC Neurology, 19*, Article 57. https://doi.org/10.1186/s12883-019-1283-9

World Health Organization. (2023). *Dementia*. https://www.who.int/news-room/fact-sheets/detail/dementia

Part 2: Empirical Paper

Breaking Cultural Barriers: Culturally fair assessment of memory and other cognitive skills within the Black community.

Abstract

Aims:

Black African/Caribbean communities in the UK experience a higher incidence of dementia and may be disadvantaged by culturally biased cognitive assessments. This study aimed to evaluate whether culture-fair cognitive screening tools offer a more culturally appropriate and equitable alternative to standard assessments for Black African/Caribbean and White British adults. It also explored participants' experiences of each assessment in terms of comfort, clarity, and cultural relevance.

Method:

A cross-sectional comparative design was used to assess cognitive performance on culture-fair and standard tools in 44 cognitively healthy adults (24 Black African/Caribbean, 20 White British). Bayesian analyses examined between- and within-group differences across matched cognitive domains. Participants completed a post-test comfort questionnaire.

Results:

Between-group comparisons showed strong evidence for no difference across most tests. Within-group analyses revealed that Black participants consistently performed better on culture-fair tests, a pattern not observed among White British participants. Feedback indicated these were experienced as more intuitive, relevant, and comfortable.

Conclusions:

The findings demonstrate that standard cognitive tests may underestimate ability in Black African/Caribbean adults, while culture-fair tools such as the CCD and RUDAS offer more equitable assessments. These tools not only reduced performance disparities but were also

rated as more culturally relevant and comfortable by participants. Clinically, this supports the need to adopt culturally fair tools to improve diagnostic accuracy and patient engagement in memory services. Policy-wise, national dementia screening guidelines should formally endorse validated culture-fair assessments to reduce systemic bias. Future research should further explore disaggregated experiences within Black communities and validate these tools in larger and more diverse UK samples.

Introduction

At present, dementia is the seventh leading cause of death among older adults globally. According to the Office for National Statistics (2023), National Records of Scotland (2023) and Northern Ireland Statistics and Research *Agency* (2023), it is the leading cause of death in the UK. There are currently estimated to be 982,000 people over 65 years of age with dementia in the UK, and it is predicted that this figure is likely to double every five years (Carnall Farrar., 2024; NHS England., 2023). The financial burden of dementia on the healthcare system and society is substantial and continues to grow, with it predicted to rise from £42 billion in 2024 to £90 billion by 2040. These statistics underscore the critical importance of prioritising early detection and diagnosis to manage the growing social and economic impact of dementia.

The benefits of receiving an early diagnosis include access to available treatment to support the management of symptoms, which can help the individual maintain a better quality of life. For the caregivers, it gives them more time to adjust to their new role, reducing the likelihood of depression or anxiety linked to increased stress (Vega & Ingram, 2013).

Research has shown that compared to the white British population, those from certain minoritised ethnic groups, particularly those of Black African or Caribbean heritage, are more likely to develop dementia (Pham et al, 2018). For example, Mukadan et al (2023) found that in England, Black African or Black Caribbean groups had 22% higher dementia incidence compared to White British groups, even though Black African and Black Caribbean groups only made up 3.4% of England's population. Evidence also suggests that the Black African or Caribbean population are less likely to receive a timely diagnosis, with these individuals often presenting with more severe symptoms of dementia when diagnosed in secondary care (Rasmussen & Langerman, 2019; Tsamakis et al, 2021).

Discrepancies between minoritised ethnic groups and their White British counterparts in accessing timely dementia diagnosis and long-term care stem from a complex interplay of systemic, cultural and social factors. Barriers include language differences, limited access to quality interpretation services, lack of cultural relevance in services that fail to account for diverse beliefs, values and practices, the stigma surrounding dementia and its intersection with ethnicity, mistrust in healthcare due to experiences of discrimination, structural biases and perceived tokenism in service provisions, limited health literacy, and a reliance on familial or community care over formal services. These factors contribute to delayed diagnosis, inadequate care and reduced help-seeking behaviours (Alzheimer's Society, 2021; Berwald et al., 2016; Giebel et al., 2021; Gove et al., 2021; Roche et al, 2020).

While previous research has explored the different interpersonal and structural issues that contribute to disparities in dementia care, limited attention has been given to the validity of cognitive screening tests used in the UK to assess cognitive impairment across culturally diverse groups. These tests assess memory, attention, language, executive function and other

cognitive domains, which are pivotal in diagnosing dementia and informing key care strategies (Grace & Morris, 2007). When interpreting these tests, various factors are taken into consideration, such as age, sex, and educational background. Clinicians are often encouraged to report the individual's cultural background or additional languages spoken, and whether this was likely to affect aspects of their performance (Donders, 2016). However, little is done to address that fundamentally the cognitive tests commonly used in UK memory clinics often fail to account for cultural differences, as they have been developed and standardised primarily on English-speaking populations in European and American contexts, leading to Eurocentric materials and normative data (Baber, 2020). Fernandez and Abe (2017) state that while basic cognitive processes are universal to all humans, these processes develop and are expressed in culturally-distinctive ways. They also state that this cultural influence affects multiple aspects of cognition, including visual perception, analytical versus holistic reasoning, spatial reasoning and even brain organisation (Carstairs et al., 2019; Seagall et al., 1996; Uzell et al., 2007). This highlights the importance of culturally fair cognitive tests, particularly for minoritised ethnic groups who already face systemic barriers to timely diagnosis and equitable care (Boakye & Mwale, 2022).

In non-specialist settings, the NICE guidelines recommend brief cognitive instruments such as the 10-point cognitive screener (10-CS) or 6-item screener (6CIT) to assess for dementia (NICE, 2018). For specialist services, the guidelines suggest the administration of further neuropsychological assessments but do not provide specific recommendations regarding which tests to use. Regarding equality and diversity, the guidelines emphasise that all information and assessments should be accessible and appropriate for diverse population groups (NICE, 2019). Commonly used specialist dementia screeners used in memory services in the UK include the Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), The Stroop test, Free and Cued Selective Reminding Test (FCSRT) or Trail Making Test (TMT) as part of their assessment.

However, these tests often present challenges for individuals whose first language is not English, those with limited literacy, or those from non-Western cultural backgrounds (Ardila, 2005; Jones, 2001; Tombaugh, 1992). For instance, Chinese and Finish population groups have shown distinct differences in memory and visual task performance on the MMSE (Salmon et al., 1989), and different language versions of the MMSE have produced widely differing cut-off scores, with optimal thresholds ranging from ≤19 to 28/29. Attempts to address these disparities through statistical adjustments have led to limited success (Cossa et al., 1997; Shim et al., 2017).

One approach that has been used to address cultural bias is the translation of tests or the use of interpreters during assessments. However, this approach carries the risk of altering the constructs being measured. For instance, when assessing language, verbal processing speed, or tests that include 'word naming task for common items', translating test materials will not maintain the same level of difficulty, test fluency or cultural relevance across different cultural groups (Fernandez & Abe., 2017; Gonthier, 2022). Zimmermann et al (2004) demonstrated that simply translating items on the FCSRT was insufficient; instead, it was necessary to select culturally relevant items to ensure the test's appropriateness for the target population. This demonstrates the extensive process required to make widely used cognitive tests culturally fair, as it necessitates addressing both cultural and semantic differences. Consequently, clinicians have few options: using cognitive tests that have already been adapted, validated, or developed for a specific language and culture or developing entirely new tests that are culturally fair across diverse groups (Boakye & Mwale, 2022). This underscores the critical need for ongoing research to identify or develop culturally fair assessment tools.

Tools developed to address cultural bias in cognitive tests include the Cross-Cultural Dementia Screening (CCD) (Delgado-Alvarez et al, 2022) and the Rowland Universal Dementia Assessment Scale (RUDAS) (Nielsen et al, 2018). The CCD, developed in the Netherlands, was validated in a study with participants of Dutch, Turkish, Moroccan or Surinamese descent. Initial performance differences between ethnic groups were observed, but these differences were eliminated after accounting for age and education differences. Similarly, the RUDAS, a brief cognitive screening tool, has been validated across several multicultural populations and has demonstrated efficacy in identifying cognitive impairments in populations with diverse cultural backgrounds. While these tools are promising, limited research in the UK has compared them to commonly used cognitive tests across different ethnic groups despite the UK's growing multicultural population.

Aim and purpose

The aim of this study is to compare cognitive screening tests designed to be 'culture-fair' (CCD and RUDAS) to standardly used tests (MMSE, FCRST, The Stroop Test, TMT) for Black African/Caribbean and White British participants. Doing so could help identify more culturally appropriate cognitive screening assessment tools.

Based on previous findings that culturally adapted tools reduce bias in cognitive screening; it was hypothesised that participants would perform better on culture-fair tests than on standard tests. Additionally, performance differences between the Black African/Caribbean groups and White British groups were expected on the standardly used tests but not on the culture-fair tests.

Given the exploratory nature of the study and the relatively small sample size, Bayesian methods were used throughout to evaluate the strength of evidence for or against the

hypothesised group differences. This approach provides a more nuanced alternative to traditional null hypothesis significance testing, as it allows for quantifying evidence in support of both the null and alternative hypotheses. Recent dementia research has adopted similar methods; for example, Requena-Komuro et al. (2022) used Bayes Factors to evaluate cognitive test performance across remote and face-to-face conditions, highlighting the value of Bayesian approaches in studies where both equivalence and difference are of interest.

Methods

Design

A cross-sectional comparative design was used to examine cognitive performance across the cognitive screening tools within and between participant groups at a single time point with no follow up.

Ethical approval and considerations

Full ethical approval was obtained from the University College London Research Committee (Ref 27829/001; Appendix C). The study was deemed high risk to participants due to sensitive information being collected (ethnic groups). All participant data were pseudonymised to protect confidentiality, and no identifiable information was shared outside the research team. Data were securely stored in accordance with the Data Protection Act (2018).

Prior to participation, prospective participants received a written information sheet (Appendix D) and provided written consent (Appendix E). They were explicitly informed that the study was not a diagnostic assessment for dementia and that participation would not indicate any future health outcomes.

Participants had the right to withdraw at any point during testing and up to 28 days after completing the tests or the point at which data analysis began if participants were recruited later, after which withdrawal was no longer feasible due to anonymisation. The primary benefit of participation was contributing to research aimed at improving culturally sensitive cognitive screening tools for dementia.

Participants

Participants eligible were cognitively healthy, aged 50 years or older, identified as White British or Black African, Black Caribbean, mixed Black heritage, resided in the UK and could communicate in English. For the purposes of the study, individuals identifying as Black African, Black Caribbean or of mixed Black heritage are collectively referred to as Black participants. Exclusion criteria included individuals who were currently or previously under investigation for neurological conditions affecting cognitive function, had concerns over their cognition, had physical difficulties that could impact test performance, such as uncorrected visual impairments or had taken a neuropsychological test within the past six months to minimise potential retest effects.

Two recruitment streams were used to ensure representation from both Black and White British participants. The Black participants were recruited through a multi-faceted approach, including social media, outreach efforts to community organisations serving predominantly Black African and Black Caribbean groups and collaboration with dementia-focused charities and organisations. Individuals expressed their interest by completing an online form, which included key information from the participant information sheet, self-screening tick boxes to confirm eligibility and a section to provide their contact details. This facilitated follow-up via

email or telephone, during which the full information sheet was shared and any further questions were addressed prior to formal enrolment in the study.

White British participants were recruited through an existing participant database maintained by the Dementia Research Centre (DRC), comprising individuals who had previously expressed interest in participating in research studies. The eligibility criteria in the DRC data pool aligned with the those of the current study, ensuring consistency across participant groups. While participants were selected with the aim of achieving comparability with Black participants on age, sex and education level, exact matching was limited by the availability of eligible White British participants.

Measures

Demographic Survey

Participants completed a brief demographic survey prior to cognitive testing. The survey was developed for the study and was not based on a pre-existing measure. It included multiple-choice and short-answer questions. The purpose was to collect background information relevant to participant matching and to support interpretation of cognitive test scores. Participants were asked to indicate their sex (male or female) and ethnicity, using categories based on the UK Government's Office for National Statistics (ONS) census classifications (e.g., Black African, Black Caribbean, Mixed with Black African, Mixed with Black Caribbean, Other Black background). Open-ended options were provided where applicable (e.g., "Other Black background: please specify") (Office for National Statistics, 2021).

Participants also reported their country of birth, year of migration to the UK (if applicable), language and country of education, highest educational qualification (e.g., Secondary School,

Undergraduate Degree, PhD) and total years of education. All responses were self-reported and completed on paper in written format.

Mini-Mental State Examination (MMSE)

The MMSE assesses orientation, registration, attention and calculation, recall, language, and visuospatial abilities (Folstein, Folstein, & McHugh, 1975). The test consists of 11 items, administered verbally and in writing, taking approximately 5–10 minutes.

Orientation was assessed through 10 questions on time (e.g., year, season, day) and place (e.g., city, building). Each correct response scores one point, with a maximum of 10 points. Registration and recall involved immediate and delayed memory tasks. Participants repeated three unrelated words (e.g., "bus, door, rose"), they scored one point per correct word. After a short delay, they recalled the words again, with a maximum of three points. Attention and calculation were assessed through two tasks. In serial sevens, participants subtracted seven from 100 up to five times (e.g., 93, 86, 79, 72, 65), they scored one point per correct response. Alternatively, they spelt "WORLD" backwards and scored one point per correctly sequenced letter. Only the higher-scoring task was counted. Language was evaluated through multiple tasks. Participants named two common objects (e.g., watch, pen), repeated the phrase "No ifs, ands, or buts," followed a three-step verbal command, read and executed a simple written instruction (e.g., "Close your eyes"), and wrote a grammatically correct sentence. Each task scored one point. Visuospatial abilities were tested by having participants copy an image of two intersecting pentagons. One point was awarded if all ten angles are present and when at least two sides intersected correctly.

Scores ranged from 0 to 30, where higher scores indicated better cognitive function. with scores of 23 or lower often indicating cognitive impairment, though cut-offs may vary based on education and cultural background (Tombaugh & McIntyre, 1992).

Studies have shown the MMSE to have high test-retest reliability (r = 0.80 – 0.89), adequate internal consistency (Cronbach's α = 0.71), and good inter-rater reliability (r = 0.75 – 0.83). For validity, it has shown moderate correlations (0.41 – 0.49) with multiple cognitive test performance (Folstein et al., 1975; Truong et al., 2024). A cut-off of 23/24 yields a sensitivity of approximately 87% and specificity of 82% for dementia detection across pooled populations in a meta-analysis of 77 studies (Mitchell, 2009). The MMSE has been evaluated in diverse populations, including African American and African-Caribbean participants (Fillenbaum et al., 1990; Manly et al., 2002). Studies have indicated that African-Caribbean individuals in the UK tend to score lower on the MMSE compared to White British individuals, even after adjusting for factors such as education and literacy (Stewart et al., 2002).

Free and Cued Selective Reminding Test (FCSRT)

The FCSRT was used to assess episodic memory and recall using controlled learning and cued retrieval (Poos et al., 2021). Participants were presented with 16 words, shown in sets of four, each belonged to a distinct semantic category. They identified each word based on its category cue before they recalled all four words. This process was repeated for all sets, with up to three trials per set until all words are recalled or trials are exhausted.

The immediate recall phase consisted of three trials where participants recalled as many words as possible within two minutes. A semantic cue was provided if a word was not recalled freely after two minutes. A 20-second interference task (counting backwards) followed each trial. After a 15–20-minute delay, the delayed recall phase mirrored the immediate recall

phase, which assessed free and cued recall. Performance on the test was assessed across four main recall scores: immediate free recall, with a maximum score of 48; immediate total recall, which included both free and cued recall, also with a maximum score of 48; delayed free recall, with a maximum score of 16; and delayed total recall, which combined free and cued recall in the delayed phase, with a maximum score of 16.

The FCSRT demonstrates good internal consistency reliability (Cronbach's $\alpha = 0.915$ for immediate recall, 0.879 for delayed recall) and also shows strong diagnostic accuracy with sensitivity of 97-99% and specificity of 94-96% for Alzheimer's Disease detection (Buschke, 1984; Grober et al., 1988; Lemos et al., 2015). The FCSRT has been validated in diverse populations, including African American participants (Grober et al., 1998).

The Stroop

The Stroop test was used to assess executive functioning, specifically cognitive interference and response inhibition (Stroop, 1935). It consisted of three conditions. In the first condition, *colour naming,* participants named the ink colour in which words are printed, with the word and ink colour being congruent (e.g., "green" written in green ink). The second condition, *word reading*, served as a baseline, where participants read words printed in black ink. In the third condition, *inhibition,* participants named the ink colour where the colour and word are incongruent (e.g., "red" written in green ink, required the response "green").

Performance was based on completion time (seconds) and the number of uncorrected and self-corrected errors across all conditions. The primary executive function measured was the interference score, calculated by subtracting the time taken in the word reading or colour naming condition from the inhibition condition. Higher interference scores indicated greater difficulty in inhibiting automatic responses. This is commonly associated with executive

dysfunction in dementia. Increased errors, particularly in the inhibition condition, may reflect attention and cognitive flexibility impairments.

Studies have shown good internal consistency reliability for the Stroop test, with Cronbach's α ranging from 0.73 – 0.89 across different conditions. Regarding validity, the Stroop test has demonstrated a sensitivity of 88% and specificity of 85%. Performance on the test can be influenced by age, education, and processing speed (Scarpina & Tagini, 2017). While normative data have been developed for African American populations in the United States (Moering et al., 2004; Norman et al., 2011), there are currently no published validation studies specifically involving Black British, Black Caribbean, or Black African populations.

Trail Making test (TMT)

The TMT was used to assess executive functioning, including selective and divided attention, visual search and scanning, processing speed, and cognitive flexibility (Reitan & Wolf, 1985). It consisted of two parts: TMT-A and TMT-B. In TMT-A, participants were required to connect 25 numbered circles in sequential order as quickly as possible. In TMT-B, participants alternated between numbers and letters in ascending order (e.g., number 1, to letter A, to number 2, to letter B) while maintaining the correct sequence.

Performance was measured by completion time (in seconds) and errors (incorrect connections or sequence errors). Longer completion times and increased errors may indicate impairments in cognitive processing speed (TMT-A) or deficits in cognitive flexibility and set-shifting (TMT-B). The TMT-B/A ratio (time taken for TMT-B divided by time taken for TMT-A) is often used to isolate executive dysfunction, as TMT-B places greater demands on cognitive flexibility and working memory. Higher TMT-B scores, particularly when disproportionately longer than TMT-A, are associated with executive dysfunction.

The TMT-A demonstrates good test-retest reliability (r = 0.76–0.90), while the TMT-B shows slightly lower reliability (r=0.70 – 0.80). Regarding construct validity, the TMT highly correlates with other tests of executive function and processing speed, with correlations ranging from 0.50 to 0.80 depending on the population (Bowie & Harvey, 2006; Mitrushina, Boone, & D'Elia, 2005). However, the TMT has not been formally validated in Black British, Black Caribbean, or Black African populations, and normative data for these groups is lacking. It is widely used to detect cognitive impairments in neurological conditions, with performance influenced by age, education, and processing speed (Tombaugh, 2004).

Rowland Universal Dementia Assessment Scale (RUDAS)

The RUDAS is a cognitive screening tool designed to minimise cultural and language biases in dementia assessment. Developed in 2004, it allows administration in a preferred language and supports interpreter use, making it particularly useful for individuals whose first language is not English (Rowland et al., 2004) . The RUDAS was used to assess memory, body orientation, praxis, visuospatial skills, judgment, and language. While conceptually similar to the Mini-Mental State Examination (Folstein et al., 1975), it has been found to be equally effective in dementia screening while being less influenced by education level (Mateos-Álvarez, Ramos-Ríos, & López-Moríñigo, 2017).

Memory was assessed through a four-word shopping list recall task, scored out of eight points. Participants recalled words immediately and after a delay, with prompted words scored as zero. Body orientation involved identifying body parts based on verbal instructions, with a maximum of five points. Praxis was tested by imitating a three-step hand movement sequence, scored from zero to two. Visuospatial skills were evaluated using a cube-drawing task, with points awarded for structural accuracy, yielding a maximum of three points. Judgment was

assessed through a scenario-based question, such as crossing a busy road, to evaluate decision-making and safety awareness. Language was measured using an animal fluency task, where participants named as many animals as possible in one minute, scoring up to eight points.

A total score of 22 or below typically indicates cognitive impairment, though cultural and linguistic factors should be considered in interpretation. The RUDAS has demonstrated strong psychometric properties across diverse cultural and linguistic groups, including Arabic-, Chinese-, Spanish- and Amharic- speaking populations, Ethiopian immigrants in Israel, and African immigrants in Australia (Rowland et al., 2004; Basic et al., 2009). It has shown strong construct validity with correlations to the MMSE ranging from 0.70 – 0.90 and has demonstrated good sensitivity (77%–89%), specificity (64%–90%), high internal consistency (Cronbach's α = 0.80) and inter-rater reliability (r = 0.99) across diverse cultural groups (Rowland et al., 2004). However, to date, there is no known validation of the RUDAS in Black British, Black Caribbean, or Sub-Saharan African populations in the UK. It has also been found to be more accurate in detecting dementia across different educational and cultural backgrounds when compared to the MMSE (Mateos-Álvarez, Ramos-Ríos, & López-Moríñigo, 2017).

Cross-Cultural Dementia Screen (CCD)

The Cross-Cultural Dementia (CCD) Screen was developed in 2005 to assess cognitive function in individuals from diverse cultural backgrounds, particularly in relation to memory, mental speed, and executive functioning (Goudsmit et al., 2017). It consists of three subtests that have been adapted from well-established cognitive measures. Two subtests, the sun-moon test and the dots test, are adaptations of the Stroop test (Stroop, 1935) and the Trail Making

Test (Reitan & Wolfson, 1985), respectively. The object memory test, which includes immediate and delayed recall components, shares similarities with the Free and Cued Selective Reminding Test (FCSRT), making the CCD a robust tool for evaluating cognitive performance across various domains.

Memory was assessed through an object memory test which involved coloured pictures of everyday items (e.g., tools, food, clothing). Participants first completed an object learning trial, followed by an immediate recognition test among an increasing number of distractor items. After a 15-minute break, they are tested again for delayed recognition. Scores were based on the number of correctly recalled target items, with higher scores indicating better memory performance. Mental speed and inhibition were measured by the sun-moon test, adapted from the Stroop test (Stroop, 1935). Participants were first asked to name pictures of suns and moons. In the second part, they had to name the opposite (e.g., say "sun" for moon), requiring inhibition of automatic responses. Performance was scored based on time taken and penalties for errors, with longer times and higher penalties indicating slower mental speed and inhibition difficulties. Divided attention was tested using the dots test, adapted from the TMT (Reitan & Wolfson, 1985). Participants connected numbered domino pieces in sequential order (one- nine). In the second part, they alternated between connecting black and white domino pieces whilst they attempted to maintain the correct order. Performance was scored by time and penalties, with increased times and penalties indicating difficulties with divided attention and processing speed.

The three subtests demonstrated high predictive validity for dementia, with 85% sensitivity and 89% specificity (Goudsmit et al., 2017). It was originally validated in the Netherlands among older adult immigrants from Turkish, Moroccan, Surinamese, and Antillean

backgrounds (Goudsmit et al., 2017). The CCD also showed correlations with the FCSRT, Stroop and TMT, supporting the convergent validity (Delgado-Álvarez et al., 2023). However, there are currently no published validation studies involving Black British, Black Caribbean, or Black African populations.

The CCD and RUDAS were grouped as Part A, representing the culture-fair test condition, while the MMSE, FCSRT, TMT and the Stroop comprised Part B, representing the standardly used test condition.

Comfort Survey

Participants completed a post-test survey developed for this study after Part A and Part B of the cognitive assessments to evaluate their experience (Appendix F). The survey included five items on overall experience, comfort, clarity of instructions, and cultural relevance of test materials, with response options ranging from "Very Good" to "Very Poor" or "Strongly Agree" to "Strongly Disagree." An open-ended question allowed additional feedback. Responses were scored based on selected categories to capture participants' comfort and perceptions.

Procedures

Testing sessions were conducted face-to-face in a private and quiet room, either at the Dementia Research Centre or in a community setting. Each session lasted between 90 minutes and 2 hours, including an optional break of up to 30 minutes between Part A (CCD and RUDAS) and Part B (MMSE, FCSRT, The Stroop and TMT) to minimise fatigue if needed. While risks were minimal, potential discomforts included the test duration and performance-related stress. Participants were offered breaks as needed and were reassured throughout that there were no right or wrong answers. Participants first completed a demographic survey followed

by cognitive tests. The test order was counterbalanced across participants to reduce order effects. The first participant started with Part A, followed by Part B, and then this alternated for subsequent participants. Cognitive tests were administered by the researcher (TO) who completed a standardised training session to ensure consistency in administering and scoring, adhering to standardised protocols. Two trained researchers (JJ and SF) from the DRC administered the tests for White British participants, also following the same training and protocols. After each part, participants completed the comfort survey. After completing the testing period, participants received a post-participation information sheet (Appendix G), were given the opportunity to ask any further questions, and received a £15 multi-store voucher for their time.

Statistical analysis

All analyses were conducted in RStudio (version 2024.12.1+563) using R and the packages *dplyr, DFBA*, *ggplot2,* and *fmsb.* Raw scores from each cognitive subtest were converted to z-scores using normed data from test manuals, which included adjustments for age, sex, and/or education where applicable. Details of the norming sources and adjustments applied for each test are summarised in Appendix H. This standardisation allowed direct comparison across tests.

Given the relatively small sample size and limitations of parametric tests in such contexts, Bayesian analysis was used throughout. The Bayesian approach was considered more appropriate than traditional frequentist methods due to its ability to quantify the strength of evidence for both the null and alternative hypotheses, rather than relying solely on significance thresholds. This was particularly useful in evaluating whether cognitive tests

performed similarly or differently across groups, where evidence of similarity can be as meaningful as evidence of difference.

In all comparisons, the null hypothesis posited that there was no meaningful difference in test performance, while the alternative hypothesis posited that there was meaningful difference. Evidence was evaluated using Bayes Factors (BF_{10}), which reflect how much more likely the observed data are under the alternative than under the null hypotheses. Specifically, $BF_{10} > 3$ indicated moderate to strong evidence for the alternative (a difference), $BF_{10} < 0.33$ indicated moderate to strong evidence for the null (no difference), and values between 0.33 and 3 were considered inconclusive. This interpretation framework is consistent with thresholds used in previous dementia research (Requena-Komuro et al., 2022). In Bayesian analysis, the posterior probability represents the degree to which the data support a given hypothesis, incorporating both observed results and any prior assumptions. While not equivalent to p-values, Bayes Factors serve a similar interpretive function by guiding conclusions about whether results support the null or alternative hypothesis (Pek & Van Zandt, 2020).

To determine the appropriate Bayesian test, the assumption of normality was assessed for each between-group and within-group comparison using the Shapiro-Wilk test. When normality was met ($p > .05$), Bayesian independent-samples t-tests (for between group) or Bayesian paired-samples t tests (for within-group) were conducted. When assumptions were violated ($p < .05$), Bayesian rank-sum tests (between-group) or Bayesian Wilcoxon signed-rank tests (within-group) were used instead. All Bayesian analyses were conducted using the *Discrete Finite Bayesian Analysis* (*DFBA)* package in RStudio.

Descriptive statistics (means and standard deviations) were calculated for all demographic characteristics, observed and z-scored test scores. Between-group comparisons evaluated

performance differences between the Black participants and White British participants across all cognitive tests. Within-group comparisons evaluated performance differences between the culture-fair tests (CCD subtests and RUDAS) and the standardly used tests (FCSRT, Stroop, TMT, and MMSE). Test pairings were selected based on conceptual alignment between test constructs and administration modalities.

- CCD Objects Immediate Recognition vs FCSRT Immediate Total Recall
- CCD Objects Delayed Recognition vs FCSRT Delayed Total Recall
- CCD Sun and Moon Naming vs Stroop Colour Naming
- CCD Sun and Moon Interference vs Stroop Interference
- CCD Dots Mental Speed vs TMT-A
- CCD Dots Divided Attention vs TMT-B
- RUDAS vs MMSE

Z-scores from FCSRT Immediate Free Recall, FCSRT Delayed Free Recall and Stroop Word Naming were excluded, as there were no directly comparable CCD subtests. These exclusions reflect key differences in test format: the standardly used tests relied on verbal recall or reading (e.g., recalling words or reading printed text), while the CCD subtests used image-based recognition or naming tasks (e.g., selecting pictures or naming visual symbols).

To aid interpretation, radar plots were used to compare group-level cognitive performance between the Black and White British participants. These plots displayed scaled mean z scores between -1.5 to 1.5 and converted to relative percentages for consistent visual comparison across domains. Paired box plots were generated to illustrate within-group differences between the culture-fair and standardly used tests.

Responses from the post-assessment comfort questionnaire were analysed descriptively to examine participants' reported experiences, comfort, clarity of instructions, and perceived cultural relevance of the test materials for both parts (Part A: culture-fair; Part B: standard). Frequencies and proportions were calculated for each categorical item, and open-ended responses were reviewed and grouped into descriptive categories based on shared content and quotes were selected to highlight key experiences.

Results

A total of 44 participants were included in the final analysis, comprising of 24 Black participants and 20 White British participants. Demographic characteristics for each group are shown in table 1. Participants ranged in age from 50 to 77 years, with the mean age of 59.46 years (SD=6.48) in the Black participant group and 68.1 years (SD=5.81) in the White British participant group. There was a slightly higher proportion of women in both groups. Years of education were broadly similar across groups, though slightly higher on average in the White British group.

Table 1. *Means (standard deviations) for Demographic Characteristics and Cognitive Test observed and z scores, by Ethnic Group.*

Measure	Black Participants (N=24)		White British Participants (n=20)	
Age (years)	59.46 (6.48)		68.1 (5.81)	
Female: Male, n (%)	16 (66.67%): 8 (33.33%)		13 (65%): 7 (35%)	
Years of Education	15.88(2.58)		16.35 (2.78)	
Culture- fair tests (/max) or (time in seconds)	Observed score M(SD)	Z score M(SD)	Observed score (M(SD)	z score M(SD)
CCD subtests:				
Objects Immediate Recognition (/122)	121.79 (0.51)	0.4 (0.25)	122 (0)	0.5 (0)
Objects Delayed Recognition (/122)	121.21 (1.35)	1.04 (0.27)	121.3 (1.13)	1.06 (0.23)
Sun and Moon Naming (s)	24.65 (6.83)	0.38 (0.49)	18.32 (3.52)	0.83 (0.25)
Sun and Moon Interference (s)	32.95 (8.82)	0.32 (0.35)	22.08 (5.54)	0.76 (0.22)
Dots Mental Speed (s)	29.12 (27.42)	0.23 (0.78)	10.5 (8.12)	0.76 (0.23)
Dots Divided Attention (s)	61.01 (38.23)	0.54 (0.48)	35.93 (16.57)	0.87 (0.21)
RUDAS (/30)	28.46 (1.44)	0.53 (0.66)	28.95 (1.19)	0.75 (0.54)
Standard tests				
FCSRT				
Immediate Free Recall (/48)	32.96 (5.03)	0.33 (1.3)	31.05 (4.57)	1.26 (2.36)
Immediate Total Recall (/48)	46.33 (2.97)	-0.14 (2.58)	47.05 (1.15)	0.76 (0.59)
Delayed Free Recall (/16)	12.58 (2.41)	0.63 (0.8)	11.65 (2.41)	-0.21 (1.72)
Delayed Total Recall (/16)	15.62 (1.24)	0.7 (3.35)	15.9 (0.31)	0.28 (0.62)
Stroop				
Colour Naming (s)	36.42 (5.38)	-0.68 (0.74)	29.6 (6.25)	0.39 (0.94)
Word Naming (s)	26.39 (5.14)	-0.46 (0.89)	21.85 (4.05)	0.47 (0.79)
Interference (s)	72.49 (28.02)	-0.22 (1.11)	55.58 (16.9)	0.77 (1.08)
TMT A (s)	43.03 (17.33)	-0.76 (0.91)	26.99 (10.56)	0.83 (1.09)
TMT B (s)	118.01 (70.09)	-1.21 (0.86)	61.94 (20.21)	0.43 (1.28)
MMSE (/30)	27.12 (2.31)	-0.22 (1.02)	29 (1.17)	0.32 (1.11)

WB, White British; M, Mean; SD, Standard Deviation; CCD subtests, Cross-Cultural Dementia screen; RUDAS, Rowland Universal Dementia Assessment Scale; FCSRT, Free and Cued Selective Reminding Test; TMT A, Trail Making Test A; TMT B, Trail Making Test B; MMSE, Mini-Mental State Examination.

Descriptive Overview of Test Performance by Ethnic Group

Mean z-scores were generally higher for the White British participants across both culture-fair and standard cognitive tests (Table 1). However, differences on the culture-fair tasks were smaller than those observed for the standard tests. In some cases, such as the CCD Objects Delayed Recognition subtest, performance was nearly equivalent across groups.

Table 3. *Bayesian Between-Group Comparisons for Cognitive Test.*

Measure	BF_{10}	Posterior Probability	Evidence
Culture-fair cognitive tests			
CCD subtests:			
Objects Immediate Recognition	**1.42×10^{-13}**	0%	Moderate to Strong evidence for null hypothesis
Objects Delayed Recognition	0.56	36.0%	Inconclusive evidence
Sun and Moon Naming	**1.47×10^{-5}**	<0.001%	Moderate to strong evidence for null hypothesis
Sun and Moon Interference	**2.36×10^{-6}**	<0.001%	Moderate to strong evidence for null hypothesis
Dots Mental Speed	**9.21×10^{-7}**	<0.001%	Moderate to strong evidence for null hypothesis
Dots Divided Attention	**0.0050**	0.5%	Moderate to strong evidence for null hypothesis
RUDAS	**0.079**	7.%	Moderate to strong evidence for null hypothesis
Standardly used cognitive tests			
FCSRT Immediate Total Recall	**0.068**	6.3%	Moderate to strong evidence for null hypothesis
FCSRT Delayed Total Recall	2.50	71.4%	Inconclusive evidence
Stroop Colour Naming	**1.14×10^{-5}**	<0.001%	Moderate to strong evidence for null hypothesis
Stroop Interference	**9.42×10^{-5}**	<0.01%	Moderate to strong evidence for null hypothesis
TMT A	**$5.63\% \times 10^{-7}$**	0%	Moderate to strong evidence for null hypothesis
TMT B	**9.29×10^{-6}**	<0.001%	Moderate to strong evidence for null hypothesis
MMSE	**0.0198**	1.9%	Moderate to strong evidence for null hypothesis

BF_{10} >3 = moderate to strong evidence for alternative hypothesis; BF_{10} <0.33= moderate to strong evidence for null hypothesis; BF_{10} 0.33 – 3 =inconclusive evidence; bold text = statistical evidence; CCD subtests, Cross-Cultural Dementia screen; RUDAS, Rowland Universal Dementia Assessment Scale; FCSRT, Free and Cued Selective Reminding Test; TMT A, Trail Making Test A; TMT B, Trail Making Test B; MMSE, Mini-Mental State Examination

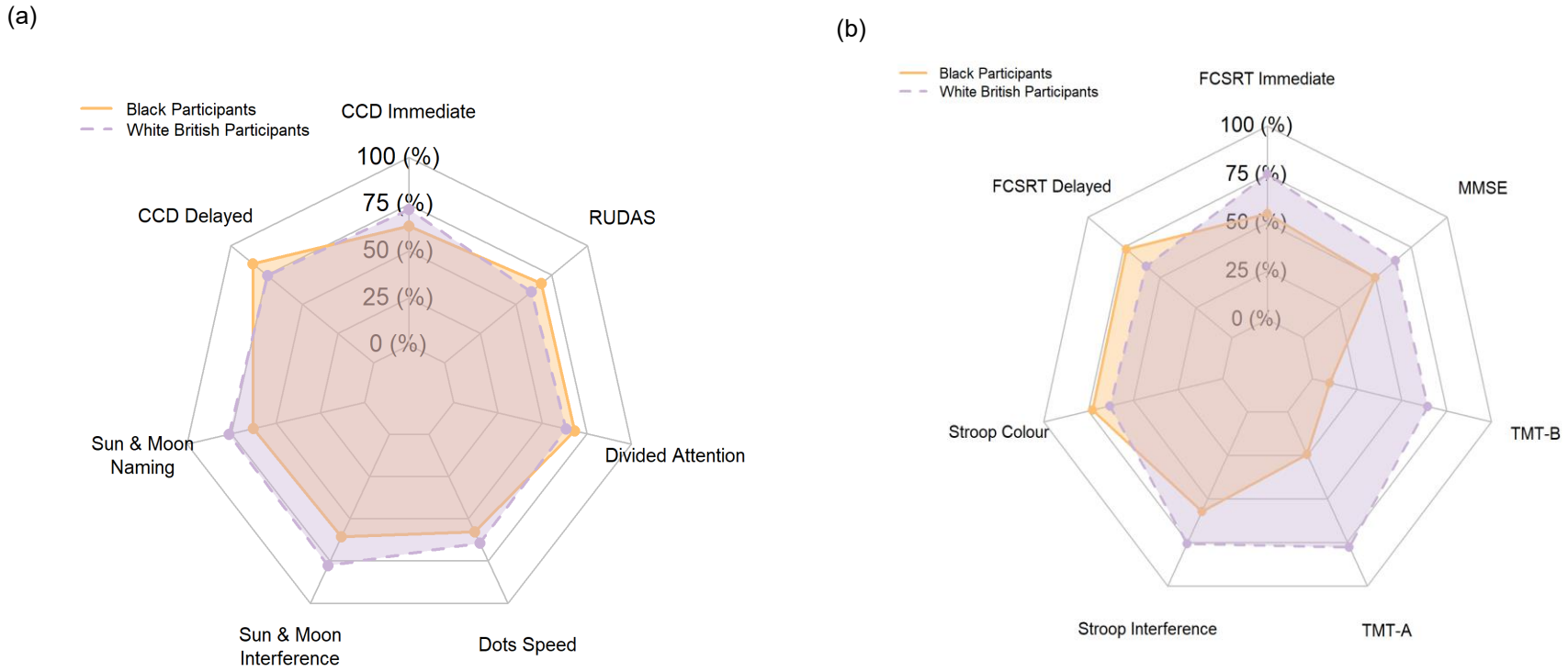


Figure 1.

Radar Plot of Performance by Cognitive Test and Ethnic Group.

Note. (a) Culture-fair tests (b) Standardly used tests Abbreviations: CCD Immediate, Cross-Cultural Dementia Objects Immediate Recognition; CCD Delayed, Cross-Cultural Dementia Objects Delayed Recognition; RUDAS, Rowland Universal Dementia Assessment Scale; Divided Attention, Cross-Cultural Dementia Dots Divided Attention; Dots Speed, Cross-Cultural Dementia Dots Mental Speed; Sun & Moon Interference, Cross-Cultural Dementia Sun and Moon Interference; Sun & Moon Naming, Cross-Cultural Dementia Sun and Moon Naming; FCSRT Immediate, Free and Cued Selective Reminding Test Immediate Total Recall; FCSRT Delayed, Free and Cued Selective Reminding Test Delayed Total Recall; MMSE, Mini-Mental State Examination; TMT A, Trail Making Test A; TMT B, Trail Making Test B; Stroop Colour, Stroop Colour Naming.

Between-Group Comparisons: Culture-fair vs Standardly Used Cognitive Tests

Minimal Differences Between the Black participants and White British Participants

There was strong evidence for no difference between the Black participants and White British participants on 12 of the 14 comparisons made (table 2). While mean z scores tended to be higher for the White British group, on these measures, the Bayesian evidence suggests that these observed differences are not statistically meaningful. The only comparisons that did not provide strong evidence for no difference were for CCD Objects Delayed Recognition and FCSRT Delayed Total Recall (BF_{10} = 2.50). The White British participants performed better on the CCD, whereas the Black participants performed better on the FCSRT; however, the evidence was inconclusive.

These statistical findings are largely supported by radar plots (Figure 1), which show overlapping gap profiles across most domains, especially on the culture-fair tests (a). However, closer visual inspection reveals some divergence in specific subtests within the standardly used tests (b). The most marked group differences were observed in TMT-A and TMT-B, where Black participant group performed notably lower than the White British group. These discrepancies were much less pronounced in the corresponding CCD Dots Mental Speed and Divided attention, suggesting that standard measures may pose greater challenges for some groups. A similar trend was seen in the FCSRT, where wider gaps appeared between groups on delayed recall tasks, despite statistical evidence being inconclusive. While the Bayesian results point to overall minimal between-group differences, the visual data highlight areas where specific standard test components may disproportionately impact certain groups.

Table 3. *Bayesian Within-Group Comparisons for Cognitive Test Pairs by Ethnic Group.*

Test pair	Black Participants			White British Participants		
	BF_{10}	Posterior probability	Evidence	BF_{10}	Posterior probability	Evidence
CCD Immediate vs FCSRT Immediate Total Recall	0.74	42.5%	Inconclusive evidence	**0.10**	9.1%	M to S evidence for null
CCD Delayed vs FCSRT Delayed Total Recall	**4363.5**	99.98%	M to S evidence for alternative	**17,092**	99.99%	M to S evidence for alternative
CCD Sun and Moon Naming vs Stroop Colour Naming	**>30,000**	100%	M to S evidence for alternative	2.06	(t test)	Inconclusive evidence
CCD Sun and Moon Interference vs Stroop Interference	**173.07**	99.4%	M to S evidence for alternative	**0.21**	17.6%	M to S evidence for null
CCD Dots Mental Speed vs TMT-A	**1490.31**	99.9%	M to S evidence for alternative	0.70	41.2%	Inconclusive evidence
CCD Dots Divided Attention vs TMT-B	**>30,000**	100%	M to S evidence for alternative	**20.29**	95.3%	M to S evidence for alternative
RUDAS vs MMSE	**204.44**	99.5%	M to S evidence for alternative	**6.80**	87.2%	M to S evidence for alternative

BF_{10} >3 = moderate to strong evidence for alternative hypothesis; BF_{10} <0.33= moderate to strong evidence for null hypothesis; BF_{10} 0.33 – 3 =inconclusive evidence; bold text = statistical evidence; CCD, Cross-Cultural Dementia screen; FCSRT, Free and Cued Selective Reminding Test; TMT A, Trail Making Test A; TMT B, Trail Making Test B; RUDAS, Rowland Universal Dementia Assessment Scale; MMSE, Mini-Mental State Examination

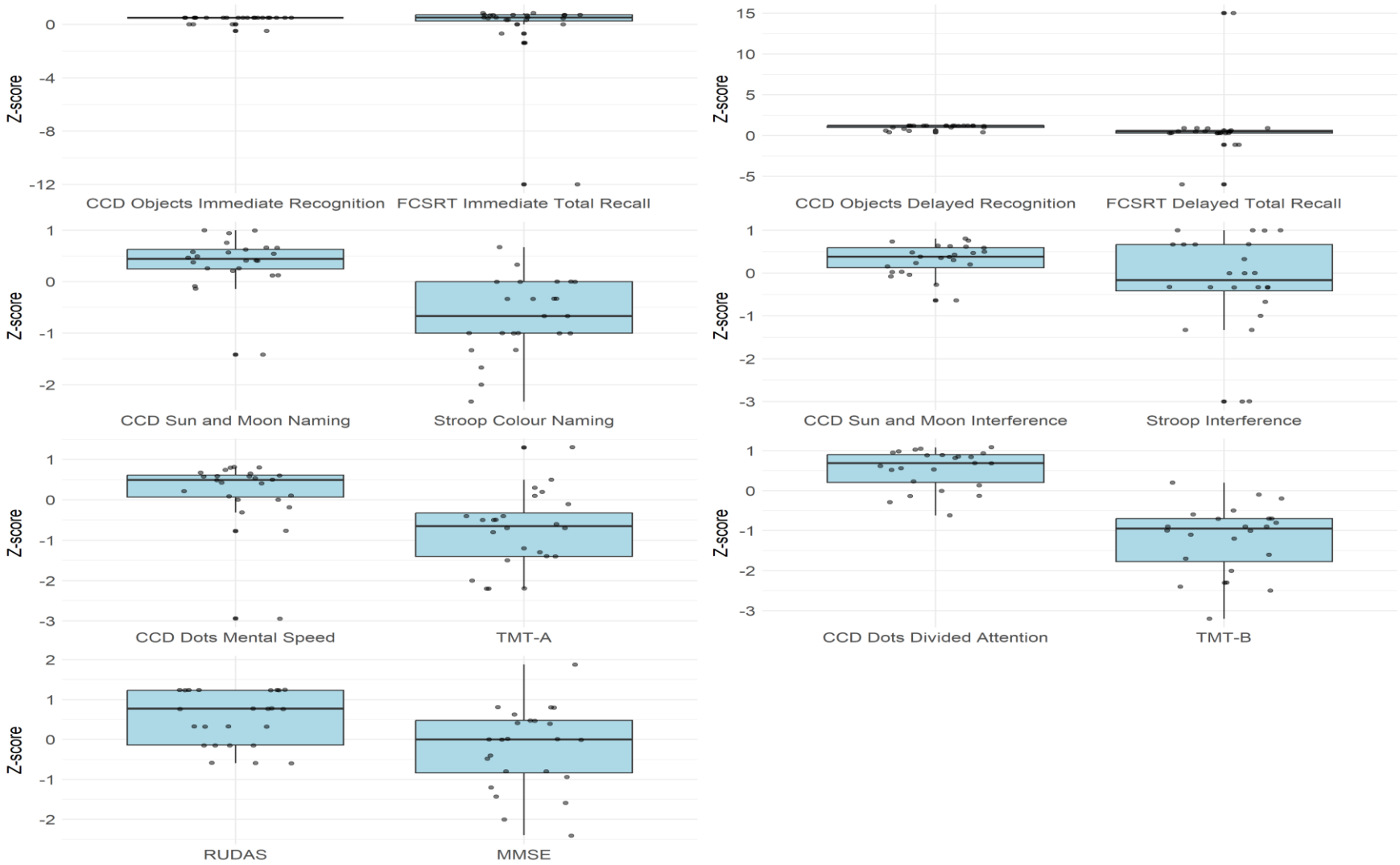


Figure 2.

Within-group comparisons of cognitive test scores among Black participants. Each panel shows paired boxplots of culture-fair vs standardly used tests.

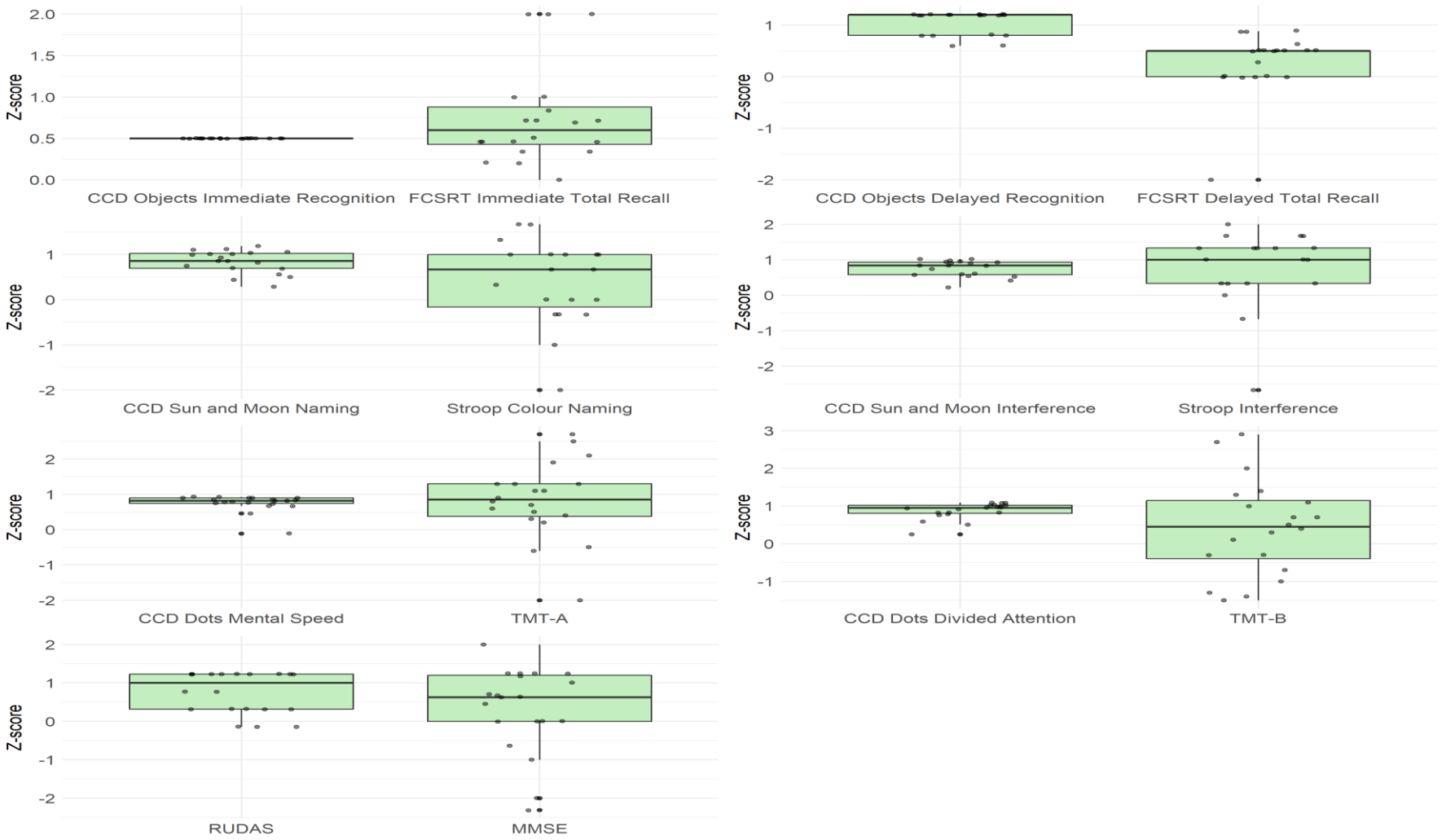


Figure 3.

Within-group comparisons of cognitive test scores among White British participants. Each panel shows paired boxplots of culture-fair vs standardly used tests.

Within-Group Comparisons: Culture-fair vs Standardly Used Cognitive Tests

Stronger Performance on Culture-fair Tests in Black Participant Group

In the Black participant group, there was moderate to strong evidence for the alternative hypothesis in all but one of the paired comparisons (6 of 7), indicating reliable performance differences between the culture-fair and standardly used cognitive tests. One comparison yielded inconclusive evidence (CCD immediate recognition vs FCSRT immediate total recall). Across test pairs, performance tended to favour the culture-fair tests, based on the mean z-scores.

These patterns are clearly reflected in the paired box plots (Figure 2), which illustrate a general trend of improved performance on culture-fair tests across most domains. The visual spread and medians indicate that Black participants generally performed better on the culture-fair tasks, reinforcing the Bayesian findings of a robust within-group performance advantage.

Mixed evidence for Culture-fair Advantage in White British Participant Group

For the White British group, findings were more mixed. Three paired comparisons showed moderate to strong evidence for the alternative hypothesis, with performance also tending to favour the culture-fair tests in these cases. However, strong evidence for no difference emerged in two comparisons (CCD immediate recognition vs FCSRT immediate total recall, and CCD sun and moon interference vs Stroop interference). Two additional comparisons (CCD sun and moon naming vs Stroop colour naming, and the CCD dots mental speed vs TMT-A) yielded inconclusive evidence, indicating that the data do not clearly support a meaningful difference in either case.

These findings are echoed in the paired box plots (Figure 3), which show more variable patterns across the test pairs in the White British group. In some comparisons, such as CCD Sun and Moon Interference vs Stroop Interference, distributions overlapped substantially, indicating minimal difference between the test pairs. While a few test pairs still showed higher median scores for the culture-fair tests, the overall visual pattern aligns with the Bayesian results, highlighting a less consistency advantage in this group compared to the Black participants.

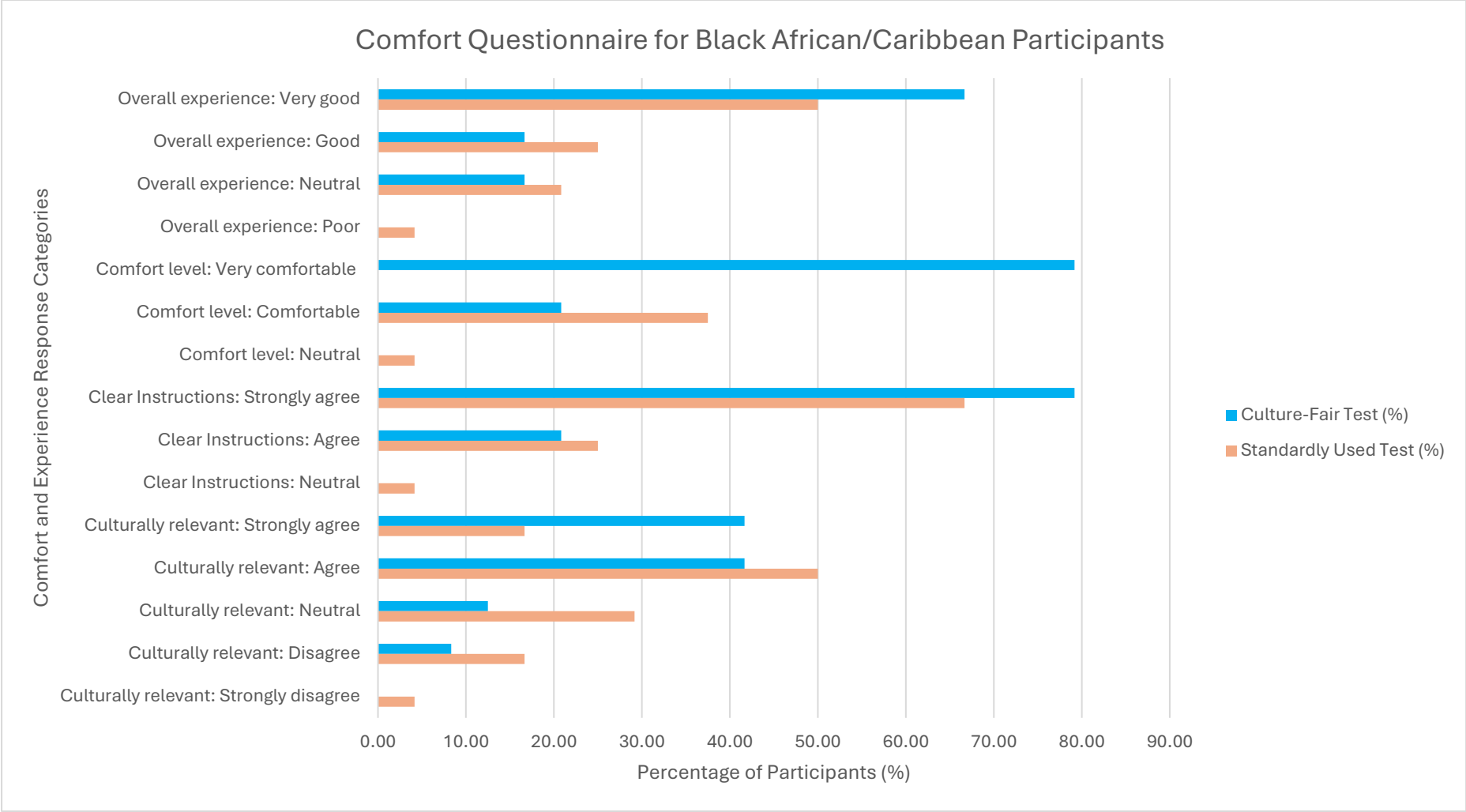


Figure 4.

Comfort and experience ratings from Black participants for culture -fair and standard tests.

Note. Categories with 0% response in both parts were excluded for clarity.

Participants Experience of Culture Fair and Standardly Used Tests

Black participants rated their experience of the culture-fair tests more favourably than the standard tests across all comfort questionnaire items. A greater proportion of participants reported a “Very good” overall experience for the culture-fair tests (67%) compared to the standard tests (50%). Comfort levels were also higher, with 79% selecting “Very comfortable” for the culture-fair test, in contrast to 38% for the standardly used test.

Clarity of instructions was more positively rated for the culture-fair tests, with 79% of participants “Strongly agreeing” that instructions were clear, compared to 67% for the standard tests. Ratings of cultural relevance also showed a preference for the culture-fair test, with 42% selecting “Strongly agree” in comparison to 17% for the standard test. Neutral and negative ratings were more common for the standard test, suggesting it was perceived as less culturally aligned.

These findings indicate that Black participants generally found the culture-fair tests to be more comfortable, clearer, and culturally appropriate than the standard cognitive screening tests.

Of the 24 Black African/Caribbean participants, 18 provided additional comments in response to the open-ended question about their experiences of completing the cognitive tests. Comments were grouped into descriptive categories based on shared content, with frequencies tallied and illustrative quotes provided to capture key experiences.

Culture-fair tests were generally perceived as intuitive and culturally aligned. The most common response was that *visual aids supported memory* (n = 3), with one participant noting: *“More visual was easier to remember; it was more difficult to focus on listening and remembering together from the other part.”* Three participants found the tasks *culturally*

familiar or relatable, such as: *"The visuals were more culturally appropriate… different fruits and globally recognisable tools and pictures."* Another three highlighted a need for *improved cultural representation*, suggesting the inclusion of items more reflective of their heritage: *"Very good experience; however, culturally appropriate fruit and veg e.g. yam, mangoes, plantain would have helped."*

Additional categories included *tasks were enjoyable or clear* (n = 2), *some tasks were challenging or unfamiliar* (n = 3), and *visuals or items felt outdated* (n = 2). Comments ranged from praise, *"Very interesting tasks"* to reflections on unfamiliar formats, *"Cube box was slightly different, as I'm not used to such task"* and criticisms of dated imagery.

In contrast, feedback on the standardly used tests reflected more difficulty and cultural mismatch. *Cultural disconnect with content* was the most endorsed category (n = 6), with one participant stating: *"I was not sure what some of the items are/were e.g. while I know daffodil is a flower, I can't identify it. I am ok with knowing a harp, because of a beer (larger) that is popular in my country, which has the harp as logo. Why not a talking drum?". Cognitive demand* was noted by four participants, particularly the strain of verbal memory tasks: *"Having to remember the words without the visuals was quite hard."* Despite this, two participants found the tests *enjoyable or stimulating*, describing them as a *"good brain teaser."*

Overall, the qualitative findings reinforced the survey data, indicating that the culture-fair tests were generally experienced as more comfortable, accessible, and culturally resonant. These insights underscore the importance of using culturally appropriate and visually supported tools in cognitive assessment with diverse populations.

Discussion

Dementia disproportionally affects people from Black African and Caribbean backgrounds in the UK, who are more likely to develop the condition and less likely to receive a timely diagnosis. While systemic and cultural barriers have been widely discussed (Berwald et al., 2016; Mukadam et al., 2022), longstanding concerns persist about the cultural inappropriateness of commonly used cognitive tests. Developed and normed in Western contexts, these tools risk underestimating ability in minoritised ethnic groups due to, linguistic or educational cultural bias (Fernandez & Abe, 2017).

While previous UK studies have acknowledged this issue, they have often examined individual tools in isolation or without validating them in diverse UK populations most focused on the MMSE, with limited exploration of alternative tests (Richards et al., 2000; Stewart et al., 2002; Rait et al., 2000; Baber, 2020). This study adds a novel dimension by comparing a broader battery of standard (MMSE, FCSRT, Stroop, TMT) and culture-fair (CCD, RUDAS) tools within the same individuals from Black African/ Caribbean and White British backgrounds. It also incorporates participants lived experiences and, most notably, applies Bayesian statistical methods. This methodological approach presents a unique contribution, enabling conclusions not only about the presence of group differences but also the strength of evidence for their absence, a nuance that traditional frequentist techniques cannot provide.

Interpretation of findings

Bayesian between-group comparisons revealed moderate to strong evidence for the null hypothesis in 12 of the 14 cognitive tests, despite descriptively higher mean z-scores in the White British group. This suggests that apparent group differences may not reflect genuine

disparities in cognitive function, but rather the cultural loading of the tests. This study builds on and extends Stewart et al. (2002), who attributed lower MMSE scores in Black African/Caribbean participants to culturally biased content, by showing that these disparities persist across other standard tools but can be diminished or disappear when using culturally fair tests. For instance, in the FCSRT Immediate Total Recall (test heavily reliant on culturally specific verbal recall) Black participants had descriptively lower scores than their White British counterparts (-0.14 vs 0.76), consistent with earlier concerns about test bias (Stewart et al., 2002). However, Bayesian analysis indicated moderate evidence for no meaningful difference between groups, suggesting that these observed disparities may not reflect true cognitive differences. In contrast, the CCD Objects Immediate Recognition, which assess the same domain (episodic memory) using visual stimuli, descriptively performance was more closely aligned between groups (0.4 vs 0.5), with the Bayes Factors supporting equivalence.

The findings support the view that while cognitive processes are universal, their behavioural expression is mediated by culture (Fernandez & Abe, 2017). Standardised tools developed and normed in Western contexts may inadvertently favour individuals whose cultural experiences align more closely with the test's assumptions. The use of Bayesian inference in this study adds a refined level of interpretation, allowing us to evaluate not only whether group differences exist, but whether the evidence supports their absence. This capacity to quantify the strength of the null hypothesis is particularly important in contexts where the goal is to evaluate the fairness of diagnostic tools across diverse populations.

This study also adds a novel aspect by conducting within-group comparisons to examine how each group performed across the standard and culture-fair tools within key cognitive domains. A consistent pattern emerged: Black African/Caribbean participants performed better on

culture-fair tests across all areas. In 6 of 7 domain-matched comparisons, Bayes Factors indicated moderate to strong evidence supporting this. For instance, participants scored significantly higher on the CCD Dots Mental Speed than on the TMT-A. Similarly, performance on the CCD Sun and Moon interference task was stronger than on the Stroop Interference, particularly for Black African/Caribbean participants, suggesting that executive functioning may be more effectively assessed using formats that are more culturally neutral. These detailed subtest-level findings illustrate the value of culture-fair assessments and go beyond general patterns seen in previous literature.

In contrast, White British participants showed fewer and smaller performance differences between standard and culture-fair tests. This asymmetry suggests that standardly used tests may be better aligned with the cultural experiences of this group, and that culture-fair tools do not disadvantage them, but rather provide a more inclusive baseline

This study was unique in combining methodological and experiential perspectives by integrating direct within-group comparisons of test performance across multiple domains with participant feedback. Together, these elements offer a more comprehensive understanding of test fairness than group comparisons alone.

This study brings a different perspective by systematically exploring how Black African/Caribbean adults experience cognitive testing, beyond just test performance, an aspect unexplored in prior research, which has typically focused on the experience of dementia within the community or diagnostic accuracy. Black African/Caribbean participants rated the culture-fair tests more positively in terms of overall experience, clarity of instructions, and cultural relevance. Qualitative feedback frequently cited the intuitive nature

of the visual tests and unfamiliarity of items in the standardly used tests such as “daffodil” or harp” in the FCSRT. Participants proposed culturally resonant alternatives like “yam” or “mangoes”, which illustrated how misalignment between test content and everyday experiences can create discomfort or disengagement.

These experiential findings align with recent calls for culturally adapted assessments (Delgado-Álvarez et al., 2023) and support research showing that lack of cultural relevance in dementia services undermines trust and delays diagnosis (Berwald et al., 2016; Giebel et al., 2021). Moreover, the positive reception of visually supported non-verbal tasks underscores the value of multimodal formats (Ardila, 2005; Carstairs et al., 2019) in diverse settings. Crucially, this study adds new, UK-based experiential evidence, highlighting not only what participants found helpful but also offering concrete examples of how current tools feel misaligned with lived experience, an area previously underexplored in the literature.

In summary, this study introduced several innovations: the use of Bayesian analysis to assess both group differences and equivalence; within-group domain-specific comparisons of standardly used and culture-fair cognitive tools; and integration of subjective feedback to evaluate the experiential validity of this tests. These contributions collectively highlight that standard tools may systematically under-represent ability in Black/African/Caribbean groups, and that culture-fair tools may offer a more equitable and inclusive alternative.

Strengths and limitations

One major strength of the study lies in its use of Bayesian statistics. This approach allows researchers to quantify evidence for both differences and the absence of meaningful differences, offering a deeper and more nuanced understanding than traditional significance

testing. This was especially valuable in the context of the small sample size, where frequentist methods may have risked type II errors. Bayesian inference allowed the study to more robustly assess whether observed group similarities likely reflect a true absence of meaningful difference rather than being attributed to random chance.

Another strength was the mixed-methods design, which strengthened the ecological validity of the findings. Participant ratings and qualitative feedback complemented performance data and confirmed that culture-fair tools were not only statistically fairer but also experienced as clearer and more engaging.

Methodological rigour was further enhanced by counterbalancing test order, reducing potential practice effects which ensured that the observed differences were not merely artefacts of repeated testing. Furthermore, community-based recruitment as opposed to relying solely on those who have participated in other dementia research studies also fostered trust among populations historically underrepresented in cognitive research and helped reduce barriers to participation among groups often marginalised by traditional recruitment methods.

Nonetheless, limitations do remain. Although all participants self-identified as either Black African/Caribbean or White British, the grouping of Black African and Caribbean individuals into a single category may have masked meaningful sub-group differences. Black communities are not monolithic; factors such as migration history, language, and or regional cultural norms differ across African and Caribbean subgroups and may have influenced both test performance and perceptions of cultural relevance. While the current sample size

precluded subgroup analyses, future research should aim to disaggregate these identities to better understand culturally specific experiences.

Although data on migration history and socioeconomic status were collected from the Black African/Caribbean participants, the limited sample size meant these variables could not be meaningfully analysed. Future research with larger and more diverse samples should prioritise examining these contextual factors, as they may offer critical insights into the interpretation of cognitive test outcomes across more diverse populations.

One further methodological limitation is that participants were not screened for colour vision deficiencies, which could have influenced performance on the Stroop test—a task dependent on distinguishing coloured stimuli. However, no participants raised concerns or showed signs during administration that suggested colour-blindness impacted their ability to complete the task.

While efforts were made to match the Black African/Caribbean and White British participants on age, sex, and education level, minor discrepancies in education and age range emerged, this was unavoidable due to recruitment constraints at the time. These minor discrepancies still could allow for potential confounding variables. Combined with the sample's relatively narrow educational range, this limits the generalisability of findings to those with lower levels of formal education, who may be more susceptible to disadvantage on language- and literacy-loaded tasks. The overall sample size was also modest, reducing statistical power; although Bayesian methods helped mitigate this, larger and more diverse samples are needed to strengthen and extend these conclusions.

Implications

These findings carry important implications for policy, clinical practice, and future research. From a policy perspective, this study reinforces the need for national dementia screening guidelines, such as those developed by NICE, to recommend culturally fair tools like the RUDAS and CCD. These tools have demonstrated greater equity in cross-cultural contexts. Addressing the use of culturally inappropriate assessments within broader health equity strategies may be a necessary step in mitigating diagnostic disparities.

Clinically, the results highlight the need for practitioners to consider the cultural relevance of cognitive assessments as part of routine practice. Culturally appropriate tools should be integrated not only for diagnostic accuracy but also to enhance patient comfort, engagement, and trust. Clinicians should be supported to interpret test performance in light of sociocultural background and to recognise the limitations of standard tools when used across cultural groups. Furthermore, staff training should continue to include awareness of cultural bias in cognitive tests and its potential impact on diagnostic outcomes.

For research, there is a clear need to disaggregate findings within the broad Black ethnic category, particularly to explore how migration history and socioeconomic status may shape test performance and experience. Larger-scale studies could investigate whether culture-fair tools contribute to earlier or more accurate dementia detection in minoritised populations. In parallel, qualitative research should continue to explore how individuals from diverse backgrounds experience cognitive assessment and what factors influence their engagement. Finally, the development and validation of cognitive screening tools should embed cultural consultation and community involvement at every stage, ensuring tools are designed with, rather than for, diverse populations. Continued research is also needed with other minoritised ethnic groups, alongside further exploration of cognitive tools developed and validated in

non-Western contexts, such as Sub-Saharan Africa to expand and strengthen the pool of culturally appropriate assessment options available in the UK.

Conclusions

The findings demonstrate that standard cognitive tests may underestimate ability in Black African/Caribbean adults, while culture-fair tools such as the CCD and RUDAS offer more equitable assessments. These tools not only reduced performance disparities but were also rated as more culturally relevant and comfortable by participants. Clinically, this supports the need to adopt culturally fair tools to improve diagnostic accuracy and patient engagement in memory services. Policy-wise, national dementia screening guidelines should formally endorse validated culture-fair assessments to reduce systemic bias. Future research should further explore disaggregated experiences within Black communities and validate these tools in larger and more diverse UK samples.

References

Alzheimer's Society. (2021). *Ethnic minority communities and dementia diagnosis*. https://www.alzheimers.org.uk/sites/default/files/2021-09/ethinic_minorities_increasing_access_to_diagnosis.pdf

Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. *Neuropsychology Review, 15*(4), 185–195. https://doi.org/10.1007/s11065-005-9180-y

Baber, Z. (2020). *An exploration of the perspectives of neuropsychologists working with clients from ethnically, culturally and linguistically diverse backgrounds* [Doctoral thesis, University of East London]. UEL Research Repository. https://doi.org/10.15123/uel.88861

Basic, D., Khoo, A., Conforti, D., Rowland, J., Vrantsidis, F., LoGiudice, D., & Prowse, R. (2009). Rowland Universal Dementia Assessment Scale, Mini-Mental State Examination and General Practitioner Assessment of Cognition in a multicultural cohort of community-dwelling older persons with early dementia. *Australian Psychologist, 44*(1), 40–53. https://doi.org/10.1080/00050060802593562

Berwald, S., Roche, M., Adelman, S., Mukadam, N., & Livingston, G. (2016). Black African and Caribbean British communities' perceptions of memory problems: "We don't do dementia." *PLOS ONE, 11*(4), e0151878. https://doi.org/10.1371/journal.pone.0151878

Boakye, N., & Mwale, A. (Eds.). (2022). *Systemic approaches to brain injury treatment: Navigating contemporary practice*. Taylor & Francis.

Bowie, C. R., & Harvey, P. D. (2006). Administration and interpretation of the Trail Making Test. *Nature Protocols, 1*(5), 2277–2281. https://doi.org/10.1038/nprot.2006.390

Buschke, H. (1984). Cued recall in amnesia. *Journal of Clinical Neuropsychology, 6*(4), 433–440. https://doi.org/10.1080/01688638408401233

Carnall Farrar. (2024, August 28). *Investing in the dementia age*. https://www.carnallfarrar.com/investing-in-the-dementia-age/

Carstairs, J. R., Myrden, A., & McIsaac, D. I. (2019). Embodiment in tests of cognitive functioning: A study of an embodied cognition task in older adults. *Dementia, 18*(6), 2192–2206. https://doi.org/10.1177/1471301217734472

Cossa, F. M., Della Sala, S., Musicco, M., & Venturi, P. (1997). Comparison of two scoring systems of the Mini-Mental State Examination as a screening test for dementia. *Journal of Clinical Epidemiology, 50*(9), 961–965. https://doi.org/10.1016/S0895-4356(97)00051-0

Delgado-Alvarez, A., Delgado-Alonso, C., Goudsmit, M., Garcia-Ramos, R., Gil-Moreno, M. J., Valles-Salgado, M., Diez-Cirarda, M., Zamarron-Cassinello, M. D., Matias-Guiu, J., & Matias-Guiu, J. A. (2023). Validation of the cross-cultural dementia screening test in Alzheimer's disease and Parkinson's disease. *Frontiers in Psychology, 13*, Article 1043721. https://doi.org/10.3389/fpsyg.2022.1043721

Delgado-Álvarez, A., Delgado-Alonso, C., Goudsmit, M., García-Ramos, R., Gil-Moreno, M., Valles-Salgado, M., Díez-Cirarda, M., Zamarrón-Cassinello, M., Matías-Guiu, J., & Matias-Guiu, J. A. (2023). Validation of the cross-cultural dementia screening test in

Alzheimer's disease and Parkinson's disease. *Frontiers in Psychology, 13*, 1043721. https://doi.org/10.3389/fpsyg.2022.1043721

Donders, J. (Ed.). (2016). *Neuropsychological report writing*. Guilford Press.

Fernández, A. L., & Abe, J. A. A. (2017). Bias in cross-cultural neuropsychological testing: Problems and possible solutions. *Culture and Brain, 6*(1), 1–35. https://doi.org/10.1007/s40167-017-0050-2

Fillenbaum, G. G., Heyman, A., Wilkinson, W. E., & Welsh, K. A. (1990). Comparison of two screening tests in Alzheimer's disease: The correlation and reliability of the Mini-Mental State Examination and the Blessed Dementia Scale. *Journal of the American Geriatrics Society, 38*(5), 492–497. https://doi.org/10.1111/j.1532-5415.1990.tb03549.x

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Giebel, C., Hanna, K., Tetlow, H., Ward, K., Shenton, J., Cannon, J., Butchard, S., Komuravelli, A., Gaughan, A., Eley, R., Rogers, C., Rajagopal, M., Limbert, S., Callaghan, S., Whittington, R., Shaw, L., & Gabbay, M. (2021). Accessing post-diagnostic dementia care before and since COVID-19 and associated inequalities: A qualitative study.

International Journal for Equity in Health, 20, 76. https://doi.org/10.1186/s12939-021-01418-1

Gonthier, C. (2022). Cross-cultural differences in visuo-spatial processing and the culture-fairness of visuo-spatial intelligence tests: An integrative review and a model for matrices tasks. *Cognitive Research: Principles and Implications, 7*(1), 1–27. https://doi.org/10.1186/s41235-022-00380-3

Goudsmit, M., Uysal-Bozkir, Ö., Parlevliet, J. L., van Campen, J. P. C. M., de Rooij, S. E., & Schmand, B. (2017). The Cross-Cultural Dementia Screening (CCD): A new neuropsychological screening instrument for dementia in elderly immigrants. *Journal of Clinical and Experimental Neuropsychology, 39*(2), 163–172. https://doi.org/10.1080/13803395.2016.1209464

Gove, D., Nielsen, T. R., Smits, C., Plejert, C., Rauf, M. A., Parveen, S., Jaakson, S., Golan-Shemesh, D., Lahav, D., Kaur, R., Herz, M. K., Monsees, J., Thyrian, J. R., & Georges, J. (2021). The challenges of achieving timely diagnosis and culturally appropriate care of people with dementia from minority ethnic groups in Europe. *International Journal of Geriatric Psychiatry, 36*(12), 1823–1828. https://doi.org/10.1002/gps.5614

Gracey, F., & Morris, R. (2007). Neuropsychological assessment. *Psychiatry, 6*(12), 499–502. https://doi.org/10.1016/j.mppsy.2007.10.005

Grober, E., Lipton, R. B., Hall, C., & Crystal, H. (1988). Memory impairment on free and cued selective reminding predicts dementia. *Neurology, 51*(2), 372–378. https://doi.org/10.1212/WNL.51.2.372

Jones, R. N., & Gallo, J. J. (2001). Education bias in the Mini-Mental State Examination. *International Psychogeriatrics, 13*(3), 299–310.

Lemos, R., Marôco, J., Simões, M. R., Santiago, B., & Santana, I. (2015). The Free and Cued Selective Reminding Test: Validation for mild cognitive impairment and Alzheimer's disease. *Journal of Neuropsychology, 9*(2), 242–257. https://doi.org/10.1111/jnp.12048

Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society, 8*(3), 341–348. https://doi.org/10.1017/S1355617702813157

Mateos-Álvarez, M., Ramos-Ríos, R., & López-Moríñigo, J. D. (2017). Comparison of the Rowland Universal Dementia Assessment Scale (RUDAS) with the Mini-Mental State Examination (MMSE) in a multicultural sample of patients with dementia. *International Journal of Geriatric Psychiatry, 32*(3), 288–295. https://doi.org/10.1002/gps.4489

Mateos-Álvarez, R., Ramos-Ríos, R., & López-Moríñigo, J. D. (2017). The Rowland Universal Dementia Assessment Scale (RUDAS): A useful screening test for dementia in multicultural populations. *International Psychogeriatrics, 29*(2), 343–345. https://doi.org/10.1017/S1041610216002047

Mitchell, A. J. (2009). A meta-analysis of the accuracy of the Mini-Mental State Examination in the detection of dementia and mild cognitive impairment. *Journal of Psychiatric Research, 43*(4), 411–431. https://doi.org/10.1016/j.jpsychires.2008.04.014

Mitrushina, M., Boone, K. B., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment* (2nd ed.). Oxford University Press.

Moering, R. G., Schinka, J. A., Mortimer, J. A., & Graves, A. B. (2004). Normative data for elderly African Americans for the Stroop Color and Word Test. *Archives of Clinical Neuropsychology, 19*(1), 61–71. https://doi.org/10.1016/S0887-6177(02)00196-6

Mukadam, N., Martson, L., Lewis, G., Mathur, R., Rait, G., & Livingston, G. (2023). Incidence, age at diagnosis and survival with dementia across ethnic groups in England: A longitudinal study using electronic health records. *Alzheimer's & Dementia, 19*(4), 1300–1307. https://doi.org/10.1002/alz.12774

National Institute for Health and Care Excellence. (2018). *Dementia: Assessment, management and support for people living with dementia and their carers* (NICE Guideline NG97). https://www.nice.org.uk/guidance/ng97/chapter/Recommendations#diagnosis

National Institute for Health and Care Excellence. (2019). *Dementia: Quality standard (QS184)*. https://www.nice.org.uk/guidance/qs184/resources/dementia-pdf-75545721373381

National Records of Scotland. (2023). *Alzheimer's and other dementia deaths*. https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/deaths

NHS England. (2023). *Dementia*. https://www.england.nhs.uk/mental-health/dementia/

Nielsen, T. R., Segers, K., Vanderaspoilden, V., Bekkhus-Wetterberg, P., Bjørkløf, G. H., Beinhoff, U., Minthon, L., Pissiota, A., Tsolaki, M., Gkioka, M., & Waldemar, G. (2018). Validation of the Rowland Universal Dementia Assessment Scale (RUDAS) in a multicultural sample across five Western European countries: Diagnostic accuracy and normative data. *International Psychogeriatrics, 31*(2), 287–296. https://doi.org/10.1017/S1041610218000832

Norman, M. A., Moore, D. J., Taylor, M., Franklin, D., Jr., Cysique, L., Ake, C., ... & Heaton, R. K. (2011). Demographically corrected norms for African Americans and Caucasians on the Hopkins Verbal Learning Test–Revised, Brief Visuospatial Memory Test–Revised, Stroop Color and Word Test, and Wisconsin Card Sorting Test 64-Card Version. *Journal of Clinical and Experimental Neuropsychology, 33*(7), 793–804. https://doi.org/10.1080/13803395.2011.559157

Northern Ireland Statistics and Research Agency. (2023). *Registrar General Annual Report 2023: Cause of death*. https://www.nisra.gov.uk/publications/registrar-general-annual-report-2023

Office for National Statistics. (2021). *Ethnic group, national identity and religion*. https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalidentityandreligion

Office for National Statistics. (2023). *Deaths registered in England and Wales: 2023*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths

Pek, J., & Van Zandt, T. (2020). Frequentist and Bayesian approaches to data analysis: Evaluation and estimation. *Psychology Learning & Teaching, 19*(1), 21–35. https://doi.org/10.1177/1475725719874542

Pham, T. M., Petersen, I., Walters, K., Raine, R., Manthorpe, J., Mukadam, N., & Cooper, C. (2018). Trends in dementia diagnosis rates in UK ethnic groups: Analysis of UK primary care data. *Clinical Epidemiology, 10*, 949–960. https://doi.org/10.2147/CLEP.S160597

Poos, J. M., Jiskoot, L. C., Bocchetta, M., Russell, L. L., Peakman, G., Convery, R. S., ... & Rohrer, J. D. (2021). Impairment of episodic memory in genetic frontotemporal dementia: A GENFI study. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 13*(1), e12185. https://doi.org/10.1002/dad2.12185

Rait, G., Burns, A., Baldwin, R., Morley, M., Chew-Graham, C., & St Leger, A. S. (2000). Validating screening instruments for cognitive impairment in older South Asians in the United Kingdom. *International Journal of Geriatric Psychiatry, 15*(1), 54–62. https://doi.org/10.1002/(SICI)1099-1166(200001)15:1

Rasmussen, J., & Langerman, H. (2019). Alzheimer's disease – Why we need early diagnosis. *Degenerative Neurological and Neuromuscular Disease, 9*, 123–128. https://doi.org/10.2147/DNND.S178804

Reitan, R. M., & Wolfson, D. (1985). *The Halstead–Reitan neuropsychological test battery: Therapy and clinical interpretation*. Neuropsychology Press.

Requena-Komuro, M.-C., Jiang, J., Dobson, L., Benhamou, E., Russell, L., Bond, R. L., Brotherhood, E. V., Greaves, C., Barker, S., Rohrer, J. D., Crutch, S. J., Warren, J. D., & Hardy, C. J. D. (2022). Remote versus face-to-face neuropsychological testing for dementia research: A comparative study in people with Alzheimer's disease, frontotemporal dementia and healthy older individuals. *BMJ Open, 12*(11), e064576. https://doi.org/10.1136/bmjopen-2022-064576

Richards, M., Brayne, C., Dening, T., Abas, M., Carter, J., Price, M., Jones, C., & Levy, R. (2000). Cognitive function in UK community-dwelling African Caribbean and white elders: A pilot study. *International Journal of Geriatric Psychiatry, 15*(7), 621–630. https://doi.org/10.1002/1099-1166(200007)15:7

Roche, M., Berwald, S., Adelman, S., Mukadam, N., & Livingston, G. (2020). *Understanding dementia in minority ethnic communities*. Race Equality Foundation. https://raceequalityfoundation.org.uk/wp-content/uploads/2020/03/Understanding-dementia-in-minority-ethnic-communities.pdf

Rowland, J. T., Storey, J. E., Basic, D., Conforti, D. A., & Dickson, H. G. (2004). The Rowland Universal Dementia Assessment Scale (RUDAS): A multicultural cognitive assessment scale. *International Psychogeriatrics, 16*(1), 13–31. https://doi.org/10.1017/S1041610204000043

Salmon, D. P., Riekkinen, P. J., Katzman, R., Zhang, M., Jin, H., & Yu, E. (1989). Cross-cultural studies of dementia: A comparison of Mini-Mental State Examination performance in Finland and China. *Archives of Neurology, 46*(7), 769–772. https://doi.org/10.1001/archneur.1989.00520430113031

Scarpina, F., & Tagini, S. (2017). The Stroop Color and Word Test. *Frontiers in Psychology, 8*, Article 557. https://doi.org/10.3389/fpsyg.2017.00557

Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The influence of culture on visual perception*. Bobbs-Merrill.

Shim, Y. S., Yang, D. W., Kim, H. J., Park, Y. H., Kim, S., & Lee, K. S. (2017). Characteristic differences in the Mini-Mental State Examination used in Asian countries. *BMC Neurology, 17*, Article 141. https://doi.org/10.1186/s12883-017-0942-1

Stewart, R., Johnson, J., Richards, M., Brayne, C., & Mann, A. (2002). The distribution of Mini-Mental State Examination scores in an older UK African-Caribbean population compared to MRC CFA study norms. *International Journal of Geriatric Psychiatry, 17*(8), 745–751. https://doi.org/10.1002/gps.677

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651

Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology, 19*(2), 203–214. https://doi.org/10.1016/S0887-6177(03)00039-8

Tombaugh, T. N., & McIntyre, N. J. (1992). The Mini-Mental State Examination: A comprehensive review. *Journal of the American Geriatrics Society, 40*(9), 922–935. https://doi.org/10.1111/j.1532-5415.1992.tb01992.x

Truong, Q. C., Cervin, M., Choo, C. C., Numbers, K., Bentvelzen, A. C., Kochan, N. A., Brodaty, H., Sachdev, P. S., & Medvedev, O. N. (2024). Examining the validity of the Mini-Mental State Examination (MMSE) and its domains using network analysis. *Psychogeriatrics, 24*(2), 259–271. https://doi.org/10.1111/psyg.13069

Tsamakis, K., Gadelrab, R., Wilson, M., Bonnici-Mallia, A, M., Hussain, L., Perera, G., Rizos, E., Das-Munshi, J., Steward, R., & Mueller, C. (2021). Dementia in people from ethnic minority backgrounds: disability, functioning and pharmacotherapy at the time of diagnosis. *Journal of American Medical Directors Association. 22 (2),* 446-452.

Uzzell, B. P., Pontón, M., & Ardila, A. (Eds.). (2007). *International handbook of cross-cultural neuropsychology*. Lawrence Erlbaum Associates.

Vega, A., & Ingram, S. (2013). *Issues in dementia caregiving: Effects on mental and physical health, intervention strategies, and research needs*. U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3774150/

Zimmermann, N., Guinle, V., Pontes, M. C., Delaere, F. J., & Fonseca, R. P. (2024). Cross-cultural linguistic adaptation and development of three alternate forms of the Free and Cued Selective Reminding Test to the Brazilian context: The role of word familiarity. *Applied Neuropsychology: Adult*. Advance online publication. https://doi.org/10.1080/23279095.2024.2415411

Part 3: Critical Appraisal

This critical appraisal comprises an exploration of the main reflections and learning points that emerged while conducting this research. It is structured around key themes that I reflected on most throughout the process including the importance of language when categorising the different types of cognitive screening tools, the influence of my positionality, challenges and insights from recruitment and community engagement, and the process of acquiring and applying new methodological skills. I also reflect on the development of my identity as a practitioner-researcher and offer ideas for future work that may build on the foundations of this study.

Learning through Language: Reflections on Terminology, Bias, and Cultural Fairness

Earlier in the research process, I initially used the term "culture-free" to describe tools such as the RUDAS and CCD. However, through a more thorough exploration of the literature, it became clear that this term is, well-meaning but misleading label, especially given the difficulty of designing tools that are truly independent of cultural influence. The term "culture-free" implies an objectivity or neutrality that is arguably unattainable, given that all cognitive tasks are embedded within linguistic, educational, and cultural assumptions to a certain degree (Melikyan et al., 2019).

In contrast, the term "culture-fair" more realistically acknowledges the effort to reduce bias while accepting that some degree of cultural influence is inevitable (Chithiramohan et al., 2024). This recognition clarified for me that even tools considered scientifically robust can still lead to unfair or misleading conclusions when used with populations whose cultural experiences differ from those the tools were originally developed for, particularly if those influences are not thoughtfully examined and addressed. This growing awareness prompted me to revise parts of literature review and empirical paper, replacing inaccurate references

to "culture-free" tools with more accurate terminology (“culture-fair”) that acknowledges the cultural context of cognitive assessments.

Researcher Reflexivity and Positionality

As a British-born woman of Nigerian heritage, with strong interest in neuropsychology, I was particularly motivated to explore the fairness of cognitive assessments in Black communities. Growing up in the UK with strong cultural ties to Nigeria gave me a dual perspective that sparked my initial interest in this topic. This part of my identity and professional interest also contributed to a deeper sense of responsibility to engage meaningfully with a community I am part of.

My background influenced the way I approached conversations with community organisations and participants, particularly around building trust and ensuring respectful engagement. Organisations made positive remarks regarding the aims of the research my cultural proximity to the community it is focused on. However, this did not necessarily translate into immediate trust or straightforward access. Building rapport still required consistent effort, patience, and transparency. Rather than seeing shared identity as a shortcut to access, I became increasingly aware of the importance of actively earning trust throughout the process.

This awareness brought with it a heightened sense of responsibility to ensure the research was conducted in a way that gave back to the community. I was conscious of the legacy of extractive research practices and did not want this project to replicate that dynamic. Offering tailored dementia-awareness workshops was one way I tried to give back, these were shaped by conversations with the organisation about what they wanted to learn about dementia and focused on questions and concerns that felt most relevant to their specific needs and interests.

These efforts were shaped by an awareness that, while I shared aspects of identity with many participants and community partners, I was also approaching this work from within an academic institution. This positioning came with its own power dynamics, and I remained mindful of how I was perceived, as both a community member and someone conducting research under institutional backing. This awareness encouraged me to be thoughtful and transparent in how I communicated and engaged throughout the process.

Embracing this dual role, researcher and community member, encouraged me to reflect more deeply on the responsibilities that come with insider positionality. As Chavez (2008) notes, positionality is not a fixed binary but a dynamic interplay of proximity and distance. Acknowledging this helped me approach the research with greater ethical awareness and humility. It also reminded me that identity-based proximity does not exempt a researcher from reflexive practice; rather, it makes such practice even more essential.

During this process, I also had the opportunity to reach out to and collaborate with the Dementia Research Centre, which provided valuable early feedback on the feasibility and clinical relevance of the study. Engaging with this institution gave me insight into how research ideas are shaped by both academic rigour and practical considerations within specialist services.

Recruitment and Community Engagement

Recruitment was one of the most interesting and, at times, challenging parts of the study. I initially assumed that community organisations would play a central role in facilitating participation, given the focus of the research. These organisations opened up very insightful and important conversations that shaped my reflections, but many were only able to support recruitment after the data collection phase had ended. In recognition of their ongoing interest

and involvement, several of these organisations expressed a desire to receive feedback on the research and participate in future dissemination activities. I plan to share the findings of the study through follow-up workshops and summary materials. This commitment to dissemination reflects an important part of my ethical stance: that research should not only extract data from communities but return something meaningful in ways that are accessible and relevant. Sharing findings in this way also supports knowledge translation and helps ensure that the communities involved can directly benefit from the insights generated.

One particularly insightful conversation occurred with a community organiser who questioned whether dementia screening could inadvertently reinforce stigma within their congregation. This prompted me to frame the dementia workshops more carefully, using strengths-based language and foregrounding stories of support and resilience. Their feedback underscored the importance of language not only in research instruments but also in community-facing aspects of the work. It also reminded me that community engagement is not a static process, but one that evolves through dialogue and requires ongoing negotiation of shared values and priorities.

The majority of participants were ultimately recruited through social media such as LinkedIn, Instagram, word of mouth including the use of whatsapp groups, advertisement from charity groups, Join Dementia Research platform and a Redeemed Fellowship Church. This discrepancy made me reflect on the different ways that trust and access operate in research. At several points, my research Instagram account was temporarily suspended, despite not breaking any of the rules or guidelines. I came to understand that algorithmic filtering may have played a role. There is growing evidence that social media platforms like Instagram disproportionately suppress content related to race, identity, and activism. Noble (2018)

argues that automated systems such as search engines are not neutral but often reinforce existing racial and gender hierarchies by deprioritising or misrepresenting content associated with marginalised communities. Building on this, Ali et al (2021) document how social media platforms have been shown to “shadow ban” users, limiting the visibility of their content without notice, particularly when they post about racial justice, Black Lives Matter, or other social issues. This experience reinforced the invisible labour often required when conducting research centring marginalised communities. I found myself having to repost content multiple times and use alternate platforms to maintain momentum, efforts that, while seemingly minor, revealed the extra work often required to ensure visibility for research focused on racial equity. The blocking of my account, in light of the literature, no longer felt like a technical glitch but part of a wider digital landscape where algorithms can reproduce the same forms of exclusion present in offline systems. These reflections strengthened my awareness of the systemic challenges faced by researchers working with marginalised communities.

While some may frame certain groups as “hard to reach,” I came to realise this framing often masks institutional shortcomings rather than community resistance. These reflections were deepened by my attendance to an inclusive research and recruitment conference following my data collection. Here, the importance of language and framing was a central theme, presenters advocated for describing such groups as “underserved” rather than “hard to reach,” shifting the focus from the perceived deficits of communities to the structural barriers they face which resonated with my experience. It was also suggested that broad ethnic categories be disaggregated to capture within-group variation more accurately. For instance, the experiences of a British-born Nigerian and a recently migrated Ethiopian participant may

differ significantly in terms of educational background, language fluency, and cultural context, all of which affect how individuals engage with research and interpret cognitive tasks.

Additionally, several conversations with community organisations led me to reflect more deeply on the internal diversity within the Black population in the UK. A number of organisations highlighted the differing histories and migration trajectories of African and Caribbean communities, particularly in terms of settlement patterns and generational ties to the UK. Caribbean communities, often described (through my conversations with different organisations during recruitment) as having longer-standing roots through the Windrush generation, may include more UK-born generations and different experiences of visibility and integration within British society. While this is not universally the case as African families also migrated during the Windrush period, the distinction was frequently drawn in discussions with different organisations as a general trend. In contrast, African communities may more frequently include first-generation migrants, whose experiences of navigating institutions, including research and health systems, can differ significantly.

These conversations prompted a deeper consideration of how historical and cultural context shapes engagement with research, including levels of trust, perceived relevance, and willingness to participate. They underscored the importance of not treating Black communities as a homogeneous group and reminded me of the value of disaggregating data in both research and policy. These insights were integrated into my discussion chapter, where I highlighted the risks of broad ethnic categories obscuring important within-group variation. In retrospect, I would have liked the study to explore differences between UK-born participants and those with migration histories more explicitly, as this could have provided greater nuance in understanding how participants engage with and interpret cognitive

assessments. These reflections will inform how I design future research, particularly with respect to intersectionality and culturally responsive methodology.

Learning Bayesian Methods: A Shift in Analytical Perspective

Before starting this project, my understanding of Bayesian statistics was minimal. I had initially planned to use traditional frequentist methods, such as t-tests and effect sizes. However, from hearing more about my external supervisor's (JJ) experience of conducting Bayesian analysis in dementia research which gave valuable insights and through further reading, I became increasingly drawn to the advantages of this approach. In particular, Bayesian analysis offered a strong fit for a research design like mine that incorporated both within-group and between-group comparisons. Unlike p-values, which can be misinterpreted or give binary conclusions, Bayesian statistics allow for degrees of belief and can incorporate prior knowledge, a quality that felt more aligned with the nuanced and context-dependent nature of this study.

One of the most compelling aspects of Bayesian analysis was its capacity to provide evidence for the null hypothesis, particularly in between-group comparisons. Whereas traditional frequentist approaches might have led me to simply report "no significant difference," the Bayesian framework allowed me to interpret these results as evidence *for similarity*, which aligned more closely with the goal of identifying equitable tools. This reframing shifted my perspective on what constitutes a "meaningful" result, and reinforced the value of methods that allow for more nuanced interpretations, especially when working across culturally diverse groups.

Learning to conduct Bayesian analyses was not easy; it required a steep learning curve, especially in terms of programming and interpretation. However, this process made me much

more engaged with the logic behind the numbers. I found myself thinking more critically about what it means to "find evidence" for or against a hypothesis, and I now feel better equipped to read and evaluate other research using Bayesian methods. More broadly, this learning journey deepened my appreciation for methodological transparency and the need to consider alternative frameworks that may be more appropriate for complex social and psychological research questions.

Practitioner Identity and Professional Development

One of the most valuable aspects of this project was the opportunity it gave me to reflect on my identity as a developing clinical psychologist and researcher. Conducting this study, required me to continuously evaluate how I was holding both roles and how they informed one another. The process reinforced my belief that effective, ethical clinical practice must be grounded in cultural awareness and responsiveness. I became more confident in trusting my clinical instincts when navigating relational dynamics in community engagement, and I learned how to integrate scientific rigour with the relational and ethical commitments I bring from clinical work.

The project also challenged me to develop new skills in areas I had less prior experience with, such as coding in R Studio. Through supervision and ongoing reflection, I became increasingly aware of areas where I wanted to grow, particularly in research design and statistical interpretation. I emerged from this process with a more confident sense of what kind of clinician-researcher I hope to be: someone able to bridge academic and community spaces, and who is committed to conducting research that is not only methodologically sound but socially relevant and empowering to the communities it involves.

These reflections also shaped my understanding of what culturally responsive neuropsychological practice might entail. In future clinical settings, I intend to bring greater attention to how assessment tools are presented, interpreted, and explained to clients and families from diverse backgrounds. I will also advocate for collaborative approaches that validate both clinical expertise and lived experience, for example, by incorporating caregiver narratives into the formulation of cognitive profiles, or by adjusting test interpretation based on educational and cultural context. This research process helped me recognise that clinical accuracy is not just a function of technical rigour, but also of cultural humility and relational sensitivity.

Future Research and Reflections

Looking ahead, this study raised several questions and avenues for future exploration. I remain interested in further examining how migration history and socioeconomic status influence cognitive assessment outcomes, particularly among culturally diverse populations. I am also keen to explore participatory approaches to research design that would involve community members earlier in the process, not only as participants but as co-developers of research priorities, tools, and dissemination strategies.

In considering participatory approaches to future research, I have become increasingly interested in models such as Community-Based Participatory Research (CBPR), which emphasise co-ownership, mutual benefit, and sustained partnerships between researchers and community members (Israel et al., 1998; Wallerstein & Duran, 2006). These frameworks challenge the extractive tendencies of traditional research and offer a more equitable way of generating knowledge. Applying such models could enhance both the relevance and uptake

of future dementia-related studies, particularly when working with communities who have historically been excluded or misrepresented in academic research

Finally, I hope to continue refining the use of Bayesian approaches in applied clinical research. I found the Bayesian lens more transparent and flexible in representing uncertainty and contextual evidence. With further training, I would like to use these methods in future projects where diverse participant groups are compared and where standard frequentist interpretations may be misleading or overly rigid.

Completing this project has not only contributed to my academic development but has strengthened my long-term commitment to pursuing research that is collaborative, culturally informed, and clinically meaningful. I carry forward the lessons learned here into the next stages of my development as both a clinician and a researcher.

Reference

Ali, S., Duguay, S., & Matamoros-Fernández, A. (2021). Performing platform governance: TikTok's responses to Black Lives Matter. *Internet Policy Review, 10*(4). https://doi.org/10.14763/2021.4.1584

Chavez, C. (2008). Conceptualizing from the inside: Advantages, complications, and demands on insider positionality. *The Qualitative Report, 13*(3), 474–494. https://doi.org/10.46743/2160-3715/2008.1589

Chithiramohan, T., Santhosh, S., Threlfall, G., Hull, L., Mukaetova-Ladinska, E. B., Subramaniam, H., & Beishon, L. (2024). Culture-fair cognitive screening tools for assessment of cognitive impairment: A systematic review. *Journal of Alzheimer's Disease Reports, 8*, 289–306. https://doi.org/10.3233/ADR-230194

Israel, B. A., Schulz, A. J., Parker, E. A., & Becker, A. B. (1998). *Review of community-based research: Assessing partnership approaches to improve public health*. *Annual Review of Public Health, 19*(1), 173–202. https://doi.org/10.1146/annurev.publhealth.19.1.173

Melikyan, Z. A., Agranovich, A. V., & Puente, A. E. (2019). Fairness in psychological testing. In G. Goldstein, D. N. Allen, & J. DeLuca (Eds.), *Handbook of Psychological Assessment* (4th ed., pp. 551–572). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-802203-0.00018-3

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Wallerstein, N., & Duran, B. (2006). *Using community-based participatory research to address health disparities*. *Health Promotion Practice, 7*(3), 312–323. https://doi.org/10.1177/1524839906289376

Appendices

Appendix A. Search Strategy for databases

Key Concept	Search Terms	Results (psycINFO)	Results (Medline)	Results (Web of Science)
1. Dementia	Exp Dementia/	97578	218315	118
2. Dementia	(Dement* or Alzheimer* or fronto temporal dement* or frontotemporal dement* or vascular dement* or lewy bod*) ab,ti. Web of science "Dement*" or "Alzheimer*" or "fronto temporal dement*" or "frontotemporal dement*" or "vascular dement*" or "lewy bod*"	132018	324311	457,366
3. Assessing cognition	(assess* or tool or cognitive screen* or cognitive tool or cognitive test or memory test or cognitive ability assessment or cognitive assessment or screening test or dementia assessment or neuropsychological assessment or neuropsychological battery or cognitive examination or cognitive ability test or cognitive impairment test) Web of science "assess*" or "tool" or "cognitive screen*" or "cognitive tool" or "cognitive test" or "memory test" or "cognitive ability assessment" or "cognitive assessment" or "screening test" or "dementia assessment" or "neuropsychological assessment" or "neuropsychological battery" or "cognitive examination" or "cognitive ability test" or "cognitive impairment test"	1099343	4994219	
4. General neuropsychological tests	Exp neuropsychological tests/	0 (error?)	196253	
5. Specific Neuropsychological tests	(ACE or Addenbrooke* or MOCA or Montreal cognitive assessment or MMSE or mini mental state examination or RUDAS or rowland universal dementia assessment scale or hopkins verbal learning test or HVLT or 6CIT or six-item cognitive impairment test or Brief Alzheimer* Screen or 7 minute screen or Test your memory or CAST or cognitive assessment screening test) Web of science "ACE" or "Addenbrooke*" or "MOCA" or "Montreal cognitive assessment" or "MMSE" or "mini mental state examination" or "RUDAS" or "rowland universal dementia assessment scale" or "hopkins verbal learning test" or "HVLT" or "6CIT" or "six-item cognitive impairment test" or "Brief Alzheimer* Screen" or "7 minute screen" or "Test your memory" or "CAST" or "cognitive assessment screening test"	69060	107896	
African countries Web of science (6-61)	6 Algeria	441	5574	
	7 Angola	225	2087	
	8 Benin	342	4979	

"Algeria" or "Angola" or "Benin" or "Botswana" or "Burkina Faso" or " Burundi" or "Cameroon" or "Cape Verde" or "Central African Republic" or "Chad" or "Comoros" or "Democratic Republic of the Congo" or "Cote d'Ivoire" or "Egypt" or "Equatorial Guinea" or "Eritrea" or "Djibouti" or "Ethiopia" or "Gabon" or "The Gambia" or "Ghana" or "Guinea" or "Guinea-Bissau" or "Kenya" or "Lesotho" or "Liberia" or "Libya" or "Madagascar" or "Malawi" or "Mali" or "Mauritania" or "Mauritius" or "Morocco" or "Mozambique" or "Namibia" or "Niger" or "Nigeria" or "Reunion" or "Rwanda" or "Saint Helena" or "Sao Tome and Principe" or "Senegal" or "Seychelles" or "Sierra Leone" or "Somali*" or "South Africa" or "Sudan" or "Swaziland" or "Tanzania" or "Togo" or "Tunisia" or "Uganda" or "Western Sahara" or "Zambia" or "Zimbabwe" or "Ethnic*" or "ethnic minority" or "Africa*" or "BAME" or "BME"	9 Botswana	967	3537	
	10 Burkina Faso	412	6112	
	11 Burundi	179	1282	
	12 Cameroon	792	10346	
	13 Cape Verde	73	656	
	14 Central African Republic	93	1429	
	15 Chad	190	1776	
	16 Comoros	13	682	
	17 Democratic Republic of the Congo	657	7353	
	18 Cote d'Ivoire	288	4830	
	19 Egypt	2639	28347	
	20 Equatorial Guinea	21	611	
	21 Eritrea	151	868	
	22 Djibouti	31	541	
	23 Ethiopia	2768	34285	
	24 Gabon	129	2546	
	25 The Gambia	197	3826	
	26 Ghana	3810	18363	
	27 Guinea	4569	166974	
	28 Guinea-Bissau	80	1451	
	29 Kenya	4522	28900	
	30 Lesotho	279	1110	
	31 Liberia	472	2390	
	32 Libya	244	2091	
	33 Madagascar	610	6605	
	34 Malawi	1400	10562	
	35 Mali	381	5267	
	36 Mauritania	62	915	
	37 Mauritius	248	1333	
	38 Morocco	947	10521	
	39 Mozambique	647	5330	
	40 Namibia	479	2486	
	41 Niger	477	16912	
	42 Nigeria	6904	48736	
	43 Reunion	1522	3996	

	44 Rwanda	1193	5153	
	45 Saint Helena	0	26	
	46 (Sao Tome and Principe)	12	229	
	47 Senegal	733	9015	
	48 Seychelles	81	970	
	49 Sierra Leone	588	3420	
	50 "Somali*"	1175	4422	
	51 South Africa	14497	68851	
	52 Sudan	745	11558	
	53 Swaziland	267	882	
	54 Tanzania	2593	20436	
	55 Togo	191	2147	
	56 Tunisia	708	12558	
	57 Uganda	3879	23501	
	58 Western Sahara	17	83	
	59 Zambia	1181	8376	
	60 Zimbabwe	1524	9382	
61 ethnic minority	61 (Ethnic* or ethnic minority or Africa* or BAME or BME).	203461	577074	
62 Total African countries and ethnic minorities	6 to 61 (OR)	237856	1007043	1,824,679
63 Assessing cognition or general neuropsychological tests or specific neuropsychological tests	3 or 4 or 5	1129143	5169057	7,707,476
1 (exp Dementia) AND 2 (Dementia) AND 63 (Assessing cognition or general neuropsychological tests or specific neuropsychological tests) AND 62 total African countries & ethnic minorities	1 AND 2 AND 63 (or) AND 62	1083	1476	2924

Appendix B. Quality Appraisal ratings

Brief breakdown of quality appraisal ratings:

CSI-D (Hall et al, 2000)

1. Content validity: **Positive rating** - Clear description of measurement aims, target population, concepts measured, and detailed item selection process involving interdisciplinary teams (*developed for community based dementia screening across different cultural settings, target population being older adults in community settings across diverse geographical areas, tools measures cognitive impairment and dementia symptoms, aligning with theoretical models of dementia."*
2. Internal consistency: Zero (0) - No Cronbach's alpha explicitly reported *although stated it was tested using it*
3. Criterion validity: **Positive rating** - ROC curves show good discriminative ability with areas >0.8, also validated against clinical dementia diagnoses at multiple international sites
4. Construct validity: **Zero (0**) - No explicit hypothesis testing reported
5. Reproducibility:
6. Agreement: **Zero (0)** - Not assessed
7. Reliability: **Positive rating** - Inter-rater reliability shows kappa ≈ 1.00 and correlations >0.99 4
8. Responsiveness: **Zero (0**) - Not evaluated
9. Floor/ceiling effects: **Zero (0)** - Not reported
10. Interpretability: **Indeterminate rating** - Mean and SD scores presented for subgroups 5, but minimal clinically important change not defined

Brief CSI-D (Prince et al., 2011)

1. Content Validity: **Positive rating** - clear description of measurement aim (dementia screening), target population (older adults in LMICs), and detailed item selection process.
2. Internal Consistency: **Indeterminate rating** - no Cronbach's alpha reported, though alternative coefficient (Loevinger H) provided
3. Criterion Validity: **Positive rating** - ROC analysis shows values well above 0.70 (AUROC 0.88-1.00)
4. Construct Validity: **0** - no information available on hypotheses testing.
5. Reproducibility (Agreement): **0** - no information on repeated measures.
6. Reproducibility (Reliability): **0** - no ICC/Kappa reported.
7. Responsiveness: **0** - no information on detecting change.
8. Floor/Ceiling Effects: **0** - no explicit reporting of lowest/highest scores.
9. Interpretability: **Negative rating** - insufficient subgroup analyses and no MIC defined.

FWT (Mormount et al, 2012)

1. Content Validity: **Positive rating** - Clear measurement aim, target population, and concepts described. Detailed description of item selection and test administration
2. Internal Consistency: No information available **(0)** - No mention of Cronbach's alpha in documents
3. Criterion Validity: **Positive rating** - Correlation with gold standard (FCSRT) exceeded 0.70
4. Construct Validity: **Positive rating** - Clear hypotheses formulated and validated
5. Reproducibility: **Negative rating** - Study explicitly acknowledges not evaluating inter-rater and intra-rater reliability
6. Responsiveness: No information available **(0)** - No data on detecting change over time

7. Floor/Ceiling Effects: **Indeterminate rating** - Some score ranges provided 8 but insufficient data on lowest/highest scores
8. Interpretability: **Positive rating** - Mean and SD scores presented for multiple subgroups. Clear cut-off values defined

ISTVF (Isaacs & Kennie, 1973)

1. Content Validity: **Positive rating** - Clear measurement aim as screening tool for dementia. Well-defined target population (elderly people). Clear concept being measured (mental function)
2. Internal Consistency: **0** (No information) - No Cronbach's alpha reported
3. Criterion Validity: **Positive rating** - Strong correlation with clinical diagnosis of dementia.
4. Construct Validity: **Indeterminate rating** - Some hypotheses tested regarding associations with physical illness and social class
5. Reproducibility (Agreement & Reliability): **0** (No information) - No ICC, Kappa, or repeated measures reported
6. Responsiveness: **0** (No information) - No data on detecting change over time
7. Floor/Ceiling Effects: **Positive rating** - Distribution of scores clearly reported
8. Interpretability: **Indeterminate rating** - Some subgroup scores presented but MIC not defined

RUDAS (Daniel et al, 2020)

1. Content Validity: **Positive rating** - clear description of measurement aim, target population, concepts, and item selection process with expert involvement
2. Internal Consistency: **Positive rating** - Cronbach's alpha = 0.73
3. Criterion Validity: **Positive rating** - ROC analysis showed AUC of 0.87
4. Construct validity: **Positive rating** - significant differences shown between diagnostic groups as hypothesised
5. Reproducibility: Reliability: **Positive rating** - ICC = 0.94 3
6. Reproducibility: Agreement: No information available **(0)**
7. Responsiveness: No information available **(0)**
8. Floor/ceiling effects: No information available **(0)**
9. Interpretability**: Positive rating** - mean and SD scores presented for multiple relevant subgroups including age, education, and diagnostic categories

Oral FCSRT (Montesinos, 2022)

1. Content Validity: **Positive rating** - Clear measurement aim, target population (illiterate adults), and expert panel validation described
2. Internal Consistency: **Positive rating** - Cronbach's alpha reported (.81 for free recall, .77 for total recall)
3. Criterion Validity: **Positive rating** - Strong correlations with gold standards (>.70) for RUDAS-PE (.85), PFAQ (.81), and CDR (.92)
4. Construct Validity: No information available **(0)** - No explicit hypothesis testing reported
5. Reproducibility Agreement: No information available **(0)**
6. Reproducibility Reliability: **Positive rating** - High ICC values (.959 for free recall, .967 for total recall)
7. Responsiveness: No information available **(0)**

8. Floor/Ceiling Effects: No information available **(0)**
9. Interpretability: **Indeterminate rating** - Only three group means/SDs presented, MIC not defined

EDQ (Arabi et al, 2016)

1. Content Validity: **Positive rating** - Clear description of measurement aim (screening early dementia), target population (elderly in primary care), and item selection process involving expert input
2. Internal Consistency: **Positive rating** - Overall Cronbach's alpha = 0.874, with subdomain alphas ranging 0.720-0.764
3. Criterion Validity: **Positive rating** - Validated against MMSE as gold standard, with sensitivity 71.2% and specificity 59.5%
4. Construct Validity: **Indeterminate rating** - Factor analysis performed but without clear pre-formulated hypotheses
5. Reproducibility (Agreement) No information available **(0)**
6. Reproducibility (Reliability): **Positive** - Good reliability with ICC = 0.764
7. Responsiveness: No information available **(0).**
8. Floor/Ceiling Effects: No information available **(0).**
9. Interpretability: No information available **(0)**

MMSE (Vissoci et al., 2019)

1. Content Validity: **Positive rating** - Clear description of measurement aim (screening early dementia), target population (elderly in primary care), and item selection process involving expert input
2. Internal Consistency: **Positive rating** - Overall Cronbach's alpha = 0.874, with subdomain alphas ranging 0.720-0.764
3. Criterion Validity: **Positive rating** - Validated against MMSE as gold standard, with sensitivity 71.2% and specificity 59.5%
4. Construct Validity: **Indeterminate rating** - Factor analysis performed but without clear pre-formulated hypotheses
5. Reproducibility (Agreement): No information reported **(0)**
6. Reproducibility (Reliability): **Positive rating** - Good reliability with ICC = 0.764.
7. Responsiveness: No information available **(0).**
8. Floor/Ceiling Effects: No information available **(0).**
9. Interpretability: No information available **(0).**

CDT

1. Content validity: **Positive rating** - Aim, target population, concepts, rationale described; limited on item involvement
2. Internal consistency: **Negative rating** - No Cronbach's alpha or related data reported
3. Criterion validity: **Positive rating** - Compared to MMSE/gold standard, ROC/AUC >0.7 reported
4. Construct validity: **0** - No formal hypotheses or structured testing done
5. Reproducibility – Agreement: **Positive rating** - Test-retest ICCs provided and acceptable

6. Reproducibility – Reliability: **Positive rating** - Inter-rater ICCs high, >0.70
7. Responsiveness: ? - No information available
8. Floor/Ceiling effect: ? - Not addressed
9. Interpretability: **Positive rating** - Means and SDs given for multiple subgroups

CERAD 10 WT

1. Content Validity: **Positive rating** - Clear measurement aim and target population described, test format and scoring clearly outlined
2. Internal Consistency: No information reported **(0) -** No Cronbach's alpha reported in the study
3. Criterion Validity: **Positive rating** - Successfully differentiated between normal and dementia groups, 83% sensitivity for WLL Sum Recall
4. Construct Validity: **Positive rating** - Hypotheses about group differences were confirmed
5. Reproducibility (Agreement): No information reported **(0)** No data on score agreement presented
6. Reproducibility (Reliability): No information reported **(0)** No ICC or Kappa values reported
7. Responsiveness: 0 (No information) - No longitudinal change data presented
8. Floor/Ceiling Effects: **Negative rating** - Floor effects noted for WLL Delayed Recall
9. Interpretability: **Positive rating -** Mean and SD scores presented for diagnostic groups, clear scoring ranges provided

GMSE (Prince et al, 2004)

1. Content Validity: **Positive rating** - Clear description provided of measurement aim (psychiatric assessment for older persons), target population (aged 60+), and concepts measured
2. Internal Consistency: **Positive rating** - Cronbach's alpha reported for multiple regions (India: 0.91, Latin America: 0.83, China: 0.88)
3. Criterion Validity: **Positive rating** - Validated against clinical diagnosis gold standard with good sensitivity in most regions
4. Construct Validity: No information available **(0)**
5. Reproducibility (Agreement): No information available (**0)**
6. Reproducibility (Reliability): No information available **(0)**
7. Responsiveness: No information available **(0)**
8. Floor/Ceiling Effects: No information available **(0)**
9. Interpretability: **Indeterminate rating** - While means and SDs are presented for some groups, MIC is not defined

Appendix C. Ethics approval letter

RESEARCH AND INNOVATION SERVICES

UCL

Dr Georgina Charlesworth
Division of Psychology and Language Sciences
Faculty of Brain Sciences
UCL

Cc: Tolulope Odumuyiwa

12 August 2024

Dear Dr Georgina Charlesworth,

Notification of Ethics Approval
Project ID/Title: 27829/001: Breaking Cultural Barriers: Culturally fair assessment of memory and other cognitive skills within the black community

Thank you for submitting the above high-risk research ethics application for review by the UCL Life and Medical Sciences Research Ethics Committee (LMS REC).

Further to your satisfactory responses to the review feedback, I am pleased to confirm in my capacity as Chair of UCL LMS REC that your study has been approved until **12 August 2025.**

Ethics approval is subject to the following conditions:

Notification of Amendments to the Research
Please seek Chair's approval for proposed amendments (to include extensions to the duration of the project) to the research for which this approval has been given. Each research project is reviewed separately and if there are significant changes to the research protocol you should seek confirmation of continued approval by completing an 'Amendment Approval Request Form' https://www.ucl.ac.uk/research-ethics/responsibilities-after-approval

Adverse Event Reporting – Serious and Non-Serious
It is your responsibility to report to the REC any unanticipated problems or adverse events involving risks to participants or others. The REC should be notified of all serious adverse events via the Research Ethics Service [redacted]) immediately after the incident occurs. Where the adverse incident is unexpected and serious, the Chair will decide whether the study should be terminated pending the opinion of an independent expert.

For non-serious adverse events, the Chair should again be notified via the Research Ethics Service within ten days of the incident occurring and provide a full written report that should include any amendments to the participant information sheet and study

protocol. The Chair will confirm that the incident is non-serious and report to the REC at the next meeting. The final view of the REC will be communicated to you.

Final Report
At the end of the data collection element of your research we ask that you submit a very brief report (1-2 paragraphs will suffice) which includes issues relating to the ethical implications of the research i.e., any issues obtaining consent, participants withdrawing from the research, confidentiality, protection of participants from physical and mental harm etc.

In addition, please:

- ensure that you follow all relevant guidance as laid out in UCL's Code of Conduct for Research;
- note that you are required to adhere to all research data/records management and storage procedures agreed as part of your application. This will be expected even after completion of the study.

With best wishes for the research.

Yours sincerely

[redacted]

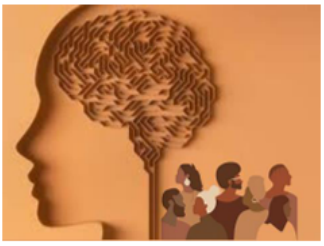
Professor Sarah Edwards
Chair, UCL Research Ethics Committee

Research Ethics Service
Research and Innovation Services
University College London
[redacted]
www.ucl.ac.uk/research-ethics/

Appendix D. Participant Information Sheet

Participant information sheet

Breaking Cultural Barriers: Culturally fair assessment of memory and other cognitive skills within the black community.



UCL Research Ethics Committee Approval ID Number: 27829/001

Principal Researcher: Dr Georgina Charlesworth

Researcher: Tolulope Odumuyiwa, Trainee Clinical Psychologist

I am a Trainee Clinical Psychologist Doctorate student at University College London (UCL). For my doctorate research project, I am collaborating with the UCL Dementia Research Centre to identify culturally fair tests that assess memory, attention, and thinking skills that are used to screen for dementia. This project is for comparing different tests and *not* designed to diagnose dementia or memory difficulties. Therefore, we hope to invite those who do **not** have any concerns about their memory or thinking skills. This will help us understand which tests are better understood across a more diverse group of people.

What is the purpose of the study?

Previous research has shown that many traditionally used memory and thinking skills tests hold *cultural bias*. This means that different aspects of the tasks, like language, pictures and concepts, are primarily based on English-speaking Europeans and North Americans from Western cultural ideals. This might include language-based tasks in English and/or pictures of things only relevant to English-speaking Europeans and North Americans.

We have identified some tests of memory and thinking skills that are designed to be more *culturally fair.* Some of these tests have been used in the Netherlands, but not with people from African or Caribbean heritage in the UK. This project therefore aims to establish if these new tests are indeed more culturally fair than tests currently used in memory clinics in the UK.

What will happen if I take part?

Participants in the study will be asked to complete a consent form and a survey collecting general demographic information this might include things like your age, sex, ethnicity, education, languages spoken, and occupation.

Taking part will involve completing Part A and Part B of memory and thinking skills tests. Each part will take no longer than one hour (a total of two hours). After completing Part A, a break will be offered. A researcher will administer the tests in person with you in an agreed quiet and private space in your community or at the UCL Dementia Research Centre (8-11 Queen Square, WC1N 3AR).

Who can take part in the study?

This study invites individuals who are 50 years or older, of Black African or Black Caribbean heritage, and living in the United Kingdom, with no concerns about their memory or thinking.Those invited to take part must be able to communicate in English, although this does not need to be their first language.

Who cannot take part in the study?

Those who are **not** able to participate in this study are individuals with uncorrected visual impairment, individuals who have participated in neuropsychological tests in the last six months and those who are under a current investigation for or have been diagnosed with a neurological condition such as dementia, stroke, Parkinson's disease or conditions following a brain injury. If you are unsure about any of the above restrictions, please contact us directly through the contact information at the end of this form.

Do I have to take part?

Participation is completely voluntary, so you do not have to participate in this study if you do not want to. If you are unsure on participating, you can ask questions about the study before you decide. If you wish to withdraw after participating, you will be asked what you would like to happen to the data you provided up to that point. It may not be possible to withdraw your data after 28 days of completing the tests, however, your results will be anonymised and you will not be identifiable.

What are the benefits of taking part?

Taking part in this study can potentially help establish culturally fair tests of memory, attention and thinking skills. This could contribute to more inclusive and accurate assessments of dementia, increasing the likelihood of individuals receiving timely support and better managing and care of dementia symptoms. Doing so is likely to have a positive impact to those living with dementia and their family and friends, reaching a wider range of communities that may otherwise be at a disadvantage. You may also find it interesting to take part to experience how memory and thinking skills are assessed. Participants will also receive a £15 multi-store voucher for their participation.

What are the potential risks of taking part?

Full ethical approval (ethics ID) has been granted by the University College London Research Ethics Committee, and the study has been deemed low risk to participants. We do not anticipate any likelihood of psychological harm or distress coming to participants. However, the testing may lead to tiredness, so the tests will be done in two parts – offering a short break from the testing and time for a drink/refreshment. If a participant were to find the testing too demanding, they would be free to withdraw from the testing period at any stage and up to 28 days after completing the tests.

What will happen to the results?

The results of the neuropsychological tests are not being used in any way to screen or diagnose dementia or any other conditions. The study results will be written into a research report and submitted as part of one of the researcher's doctoral thesis. It will also be put forward to be published in scientific journals and disseminated across relevant platforms. Please be assured that your participation will not be identifiable in any publications and all data will remain anonymised. If you opt to receive a summary of the study findings it will be made available to you when available.

What will happen to my data?

This is a collaborative study with UCL Dementia Research Centre (DRC) which is a hub for research involving people with and without dementia. As part of this collaboration, anonymised research data will be shared with the DRC for future analysis as part of their research. The DRC is also offering you the opportunity to participate in another dementia research study, further details will be provided if you would like to take part.

Taking part is completely confidential and your personal data will be managed with care and in accordance with the relevant data protection regulations (*Data Protection Action 2018 and General Data Protection Regulation*). We will retain some personal identifiable data e.g. your contact details and the signed consent form. Where possible we will pseudonymise your data (i.e., we will remove your personal information and replace it with a code that is only known to the research team). Your data will be held in a highly secure online portal called the 'UCL Data Safe Haven' and therefore protected from unauthorised access and will only be held for as long as it is required for the research.

University College London (UCL) is the data controller for this study that determines how your personal data is used in this study. Further information on how UCL uses participant information can be found here https://www.ucl.ac.uk/legal-services/privacy/ucl-general-privacy-notice-participants-and-researchers-health-and-care-research-studies. The UCL Data Protection Officer can be contacted at data-protection@ucl.ac.uk .

Who can I contact if I have questions?

If you have further questions or concerns about the study in any way, you can contact the following people below:

Dr Georgina Charlesworth (Principal Researcher)

Email: [redacted]

Tolulope Odumuyiwa (Trainee Clinical Psychologist)

Email: [redacted]

Dr Jess Jiang (External Supervisor – UCL Dementia Research Centre)

Email: [redacted]

Thank you for reading this information sheet and considering participating in this study.

Appendix E. Consent Form

Consent Form

Breaking Cultural Barriers: Culturally fair assessment of memory and other cognitive skills within the black community.

Thank you for considering taking part in this study. If you have any questions from the Information Sheet or from further information provided, please ask the researcher before you decide whether to participate. You will be given a copy of this Consent Form to keep and refer to at any time. **Please complete this form after you have read the Participant Information Sheet**

Researcher name: Tolulope Odumuyiwa, Trainee Clinical Psychologist

Email: [redacted]

Principal researcher: Dr Georgina Charlesworth

Email: [redacted]

UCL Data Protection Officer email: [redacted]

UCL Research Ethics Committee Approval ID number: 27829/001

	Please initial Boxes
I confirm that I have read and understood the Information Sheet for the above study. I have had the opportunity to consider the information and what will be expected of me. I have also had the opportunity to ask questions which have been answered to my satisfaction.	
I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason. I understand that I will be able to withdraw my data up until 28 days from today.	
I consent to participate in the study. I understand that my personal information (age, sex, ethnicity, country of birth, year of migration to the UK (if relevant), level and years of education (language educated in), occupation and post-code) will be used for the purposes explained to me.	
I understand that all personal information will remain confidential and that all efforts will be made to ensure I cannot be identified. I understand that my data gathered in this study will be stored securely. It will not be possible to identify me in any publications.	
I understand that I will not receive any individual results from the study.	
I understand that this study is in collaboration with the Dementia Research Centre.	
I understand that anonymised research data will be shared with the Dementia Research Centre as part of their collaboration with this study.	
I am aware of who I should contact if I wish to lodge a complaint	
I wish to receive a summary of the study findings.	Yes / No

Participant name and Signature: _____

Date: _____

Researcher name and signature: Tolulope Odumuyiwa [redacted]

Date: _____

Appendix F. Comfort Survey

Post Memory and Thinking Skills Test Survey

Participant ID number: _____

Part A

1. How would you rate your overall experience during the memory and thinking skills tests?

Very Good / Good / Neutral / Poor / Very Poor

2. How comfortable did you feel while completing the tests?

Very Comfortable / Comfortable / Neutral / Uncomfortable / Very Uncomfortable

3. Did you find the instructions for the tests clear and easy to understand?

Strongly Agree / Agree / Neutral / Disagree / Strongly Disagree

4. Did you feel that the test materials (e.g., vocabulary, pictures, concepts) were culturally relevant to you?

Strongly Agree / Agree / Neutral / Disagree / Strongly Disagree

5. Is there anything else would like to share about your experience of the memory and thinking skills test?

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Part B

1. How would you rate your overall experience during the memory and thinking skills tests?

Very Good / Good / Neutral / Poor / Very Poor

2. How comfortable did you feel while completing the tests?

Very Comfortable / Comfortable / Neutral / Uncomfortable / Very Uncomfortable

3. Did you find the instructions for the tests clear and easy to understand?

Strongly Agree / Agree / Neutral / Disagree / Strongly Disagree

4. Did you feel that the test materials (e.g., vocabulary, pictures, concepts) were culturally relevant to you?

Strongly Agree / Agree / Neutral / Disagree / Strongly Disagree

5. Is there anything else would like to share about your experience of the memory and thinking skills test?

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Appendix G. Post-Participant Information Sheet

Post-participant information sheet

Breaking Cultural Barriers: Culturally fair assessment of memory and other cognitive skills within the black community.

Thank you for taking part in this study. We hope that you have enjoyed taking part in the memory and thinking skills tests. Your performance in the study is not used in any way to screen or diagnose dementia or any other conditions but used to compare the different tests to each other. There is a possibility that going through the memory and thinking skills tests has got you thinking about others you may know to be experiencing difficulties in these areas. If you would like further information on dementia organisations or psychological support, the following organisations are able to offer information and advice.

Dementia organisations

- https://www.nhs.uk/conditions/dementia/
- Alzheimer's society website https://www.alzheimers.org.uk/?gad_source=1&gclid=Cj0KCQjw6auyBhDzARIsALIo6v_jJVe88eOyh1Z10ie2bFwm3AMR9jZ1M3dEG9b_0ZscJ3KwWoY9uRcaAn38EALw_wcB&gclsrc=aw.ds or dementia support line 0333 150 3456

Psychological support

- Improving Access to Psychological Therapies (IAPT) "Talking Therapies" service for those feeling worried, low or stressed. Follow the link to find your local service to self-refer or you can ask your GP to refer you https://www.nhs.uk/service-search/mental-health/find-an-nhs-talking-therapies-service
- The Black, African and Asian Therapy Network https://www.baatn.org.uk/

If you have any further questions/comments following your participation, please contact Tolu Odumuyiwa via [redacted] or the Principal Researcher, Dr Georgina Charlesworth via [redacted]

Thanks once again for your time in participating.

Appendix H. Cognitive Test Norms and Adjustments

Cognitive Test	**Norming Variables Used**
FCSRT subtests	Age, Sex
Stroop	Age
TMT (A&B)	Age, Sex, Education
MMSE	Age Education
RUDAS	Education
CCD subtests	Total sample norms (not group-specific