



OPEN ACCESS

EDITED BY

Annalena Venneri,
Brunel University London, United Kingdom

REVIEWED BY

Michael Malek-Ahmadi,
Banner Alzheimer's Institute, United States
Cândida Alves,
Federal University of Maranhão, Brazil

*CORRESPONDENCE

Kristen E. Kehl-Floberg
✉ kekehl@wisc.edu

RECEIVED 23 April 2023

ACCEPTED 08 August 2023

PUBLISHED 29 August 2023

CITATION

Kehl-Floberg KE, Marks TS, Edwards DF and
Giles GM (2023) Conventional clock drawing
tests have low to moderate reliability
and validity for detecting subtle cognitive
impairments in community-dwelling older
adults.

Front. Aging Neurosci. 15:1210585.
doi: 10.3389/fnagi.2023.1210585

COPYRIGHT

© 2023 Kehl-Floberg, Marks, Edwards and
Giles. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction is
permitted which does not comply with
these terms.

Conventional clock drawing tests have low to moderate reliability and validity for detecting subtle cognitive impairments in community-dwelling older adults

Kristen E. Kehl-Floberg^{1*}, Timothy S. Marks²,
Dorothy F. Edwards^{1,2} and Gordon M. Giles³

¹Institute for Clinical and Translational Science, University of Wisconsin-Madison, Madison, WI, United States, ²Department of Kinesiology-Occupational Therapy, University of Wisconsin-Madison, Madison, WI, United States, ³Department of Occupational Therapy, Samuel Merritt University, Oakland, CA, United States

Background: Early identification of subtle cognitive decline in community-dwelling older adults is critical, as mild cognitive impairment contributes to disability and can be a precursor to dementia. The clock drawing test (CDT) is a widely adopted cognitive screening measure for dementia, however, the reliability and validity of paper-and-pencil CDT scoring scales for mild cognitive impairment in community samples of older adults is less well established. We examined the reliability, sensitivity and specificity, and construct validity of two free-drawn clock drawing test scales—the Rouleau System and the Clock Drawing Interpretation Scale (CDIS)—for subtle cognitive decline in community-dwelling older adults.

Methods: We analyzed Rouleau and CDIS scores of 310 community-dwelling older adults who had MoCA scores of 20 or above. For each scale we computed Cronbach's alpha, receiver operating characteristic curves (ROC) for sensitivity and specificity using the MoCA as the index measure, and item response theory models for difficulty level.

Results: Our sample was 75% female and 85% Caucasian with a mean education of 16 years. The Rouleau scale had excellent interrater reliability (94%), poor internal consistency [0.37 (0.48)], low sensitivity (0.59) and moderate specificity (0.71) at a score of 9. The CDIS scale had good interrater reliability (88%), moderate internal consistency [0.66 (0.09)], moderate sensitivity (0.78) and low specificity (0.45) at a score of 19. In the item response models, both scales' total scores gave the most information at lower cognitive levels.

Conclusion: In our community-dwelling sample, the CDIS's psychometric properties were better in most respects than the Rouleau for use as a screening

instrument. Both scales provide valuable information to clinicians screening older adults for cognitive change, but should be interpreted in the setting of a global cognitive battery and not as stand-alone instruments.

KEYWORDS

clock drawing, reliability, validity, cognitive screening, older adults, community sample, mild cognitive impairment

1. Introduction

Mild cognitive impairment (MCI) is estimated to affect about 7% of adults aged 60–65, increasing to 25% by age 80–85 years (Petersen et al., 2018). MCI is defined by the American Psychiatric Association as subtle decline from baseline cognition accompanied by decreased performance of cognitively-complex daily activities (American Psychiatric Association [APA], 2013; Petersen et al., 2018). Evidence guiding clinical screening of cognition during routine health care for healthy older adults is evolving along with recent advances in the understanding of both normal aging and neurocognitive disease (Livingston et al., 2020). Recognition has grown that healthy older adults may benefit from cognitive screening as part of preventative health care (Anderson, 2019) because cognitive impairment contributes to disability (Jekel et al., 2015; Lindbergh et al., 2016) and has been found to progress into dementia in about 15% of cases (Petersen et al., 2018). In clinical practice, accurate early identification of mildly-impaired cognition enables assessment and treatment of causes of reversible cognitive decline (e.g., sleep or mood disorders), pursuit of non-pharmacological interventions to support cognitive aging (Tsolaki et al., 2011), and management of comorbidities that contribute to dementia risk (Livingston et al., 2020). Thus there is a critical need for rapid, reliable, and sensitive objective screening tools that detect subtle cognitive change from baseline (Anderson, 2019).

However, to meet these evolving clinical needs, more evidence is needed on the value of cognitive screening in healthy older adults. The U.S. Preventive Services Task Force rates the evidence for benefit from screening asymptomatic older adults for cognitive decline as insufficient (Lin et al., 2013), and many cognitive screening instruments may not be reliable for this purpose as they were designed to detect dementia, not subtle impairment. The clock drawing test (CDT) is one such classic cognitive screening instrument. The CDT is a paper-and-pencil tool developed for dementia screening, administered by asking the patient to draw an analog clock face with a specific time setting which is then scored by the assessor using a scale of criteria for success. Although developed over one hundred years ago (Hazan et al., 2018), it has recently gained interest for its potential to detect mild impairment (Duro et al., 2019), its neurofunctional correlations with the executive functions required to complete complex daily activities (Royall et al., 1999; Dion et al., 2022), and an adaptation to digital format (Müller et al., 2017; Davoudi et al., 2020; Dion et al., 2020; Yuan et al., 2021).

Clock drawing is of interest for screening healthy older adults based on its plausible neurocognitive relevance to

milder impairment (Strauss et al., 2006) such as its reliance on constructional praxis and executive function and their neuroanatomical correlates (Talwar et al., 2019) which are often affected in early dementia and MCI (Mendez et al., 1992; Rouleau et al., 1992, 1996; Chiu et al., 2008; Talwar et al., 2019). A free-drawn electronic CDT using an instrumented pen and electronic tablet-style writing surface has been found to distinguish people with MCI from healthy controls (Yuan et al., 2021), supporting the assumption that the cognitive skills required to draw a clock are relevant to the diagnostic purpose of screening for mild impairment. Compared to this electronic version, paper versions are not as accurate in detecting subtle change (Müller et al., 2017); however, conventional versions are more practical, brief, easy to obtain and administer, familiar [a CDT task is incorporated into several test batteries such as the Mini-Cog (Lam et al., 2011) and the Montreal Cognitive Assessment (Nasreddine et al., 2005)], and still have potential to yield rich clinical information. This makes conventional CDT appealing as a screening instrument for mild impairment in community-dwelling older adults.

Despite its benefits and potential, further examination of conventional CDT's appropriateness for detection of subtle impairment is needed. First, some of the index measures used in initial validation studies have since been shown to be insufficiently sensitive to MCI in a range of patient populations (Arevalo-Rodriguez et al., 2021), so these psychometric references do not relate to current diagnostic practice or evidence. Second, it is unknown whether the scoring ranges for many versions, scaled to detect potential dementia, can detect small but meaningful decreases from an unimpaired cognitive baseline. Third, although paper-and-pencil CDT's have been shown to have low sensitivity to MCI (Pinto and Peters, 2009; Duro et al., 2019), only a few recent studies have focused on community-dwelling older adults (one of the most important populations to screen for MCI). Particularly in older studies establishing cut scores, study design precluded conclusions about mild impairment. Finally, we were not able to find previous work that used methods such as item response theory to determine whether these scales' levels of difficulty are suitable for screening this population. Although item response theory has been employed to evaluate tests that include a clock drawing task [e.g., the Montreal Cognitive Assessment (Luo et al., 2020), the Texas Functional Living Scale (Lowe and Linck, 2021)] its application to stand-alone clock scales has not been reported. Because (a) the CDT was developed primarily to detect dementia, (b) the diagnostic scope of neurocognitive disorder has since expanded to include MCI, (c) gold standard screening tools are now more sensitive

to MCI, and (d) more evidence is needed for community-dwelling older adults who are the target population for screening, further study is needed to confirm conventional clock drawing's psychometric properties for detecting small differences between unimpaired and mildly impaired cognition in community-dwelling older adults.

This study examined the reliability, sensitivity and specificity, and construct validity of two CDT scoring systems, the Rouleau (Rouleau et al., 1992, 1996) and the Clock Drawing Interpretation Scale (CDIS) (Mendez et al., 1992), for identifying cognitive performance consistent with MCI in community-dwelling older adults. We hypothesized that (a) the CDIS would exhibit higher reliability and validity than the Rouleau, and (b) both scales would distinguish people with subtle cognitive deficits from cognitively unimpaired individuals at a score close to the maximum. Additionally, we will examine the construct validity of both scales' measurement of cognitive ability.

2. Materials and methods

2.1. Design

This cross-sectional observational study was approved by the University of Wisconsin-Madison Institutional Review Board. All participants provided written informed consent.

2.2. Participants

A convenience sample of $N = 345$ community-dwelling older adults was recruited in and around Madison, Wisconsin through flyers posted in community spaces, in-person recruitment at community events, and word-of-mouth. Inclusion criteria were age 55 years and older, living in the community, self-reported independence with daily activities, and ability to read and write in English.

2.3. Instruments

To obtain clock drawings, participants were instructed to draw a clock free-hand with the time setting "ten past eleven." These drawings were scored on the Rouleau and CDIS scales by two licensed and registered occupational therapists (TM, GG) and two occupational therapy graduate students under their supervision. These scores were combined into one data set, and identifiers linking scores with raters were removed.

2.3.1. Criterion measure: the Montreal Cognitive Assessment

The Montreal Cognitive Assessment (MoCA) is a brief screening of cognition (Nasreddine et al., 2005). It is a 10-min, paper-and-pencil screen of global cognition, encompassing multiple neurocognitive domains. It is scored from 0 to 30. For broad use, score ranges are designated at 30–26 for unimpaired cognition, 25–20 for performance consistent with mild cognitive impairment, and 0–19 consistent with dementia

(Nasreddine et al., 2005); however, alternative cut scores of 24 (Townley et al., 2019) to 23 (Carson et al., 2018) have been proposed for community-dwelling older adults and other specific populations (Milani et al., 2018; Rossetti et al., 2019; Townley et al., 2019). Compared to the Mini Mental Status Exam (the criterion measure in the original validation papers for these clock scales), the MoCA has superior sensitivity to MCI (90% for MoCA versus 18% for the MMSE) and good positive and negative predictive values (89 and 91%, respectively). As the criterion measure for the present study, the MoCA has the added advantage of a CDT task with the same "face, numbers, hands" structure as the Rouleau and CDIS.

2.3.2. Clock drawing test scales: the Rouleau and the Clock Drawing Interpretation Scale

We selected these scales for several characteristics that maximize their suitability for mild impairment screening. First, they have similar structures; the scoring criteria are all sequentially organized by the clock face, numbers, and hands. The relative contributions of each of these aspects is only partially understood, however, errors in hand placement have been suggested as indication of need for further assessment (Esteban-Santillan et al., 1998). Second, their difficulty is among the highest of the CDT versions because the clocks are drawn free-hand using a three-step command, which has been found to induce more errors than copied versions (Chiu et al., 2008). This suggests a more difficult task, which may be more likely to discriminate between persons with and without mild impairment. Third, they are among the largest and most detailed item banks of all the paper-and-pencil versions, increasing their potential reliability (Bandalos, 2018a). Finally, the clinical utility of this type of scale—using an unadjusted score based on criteria that emphasize qualitative aspects of performance—have been found to have the best clinical utility (Shulman, 2000). See **Appendix A** for score criteria.

2.3.2.1. Rouleau clock drawing test scale

Rouleau's scale is a modified version of a scale published by Sunderland et al. (1989); Rouleau et al. (1992). It is scored from 0 (most impaired) to 10 (least impaired), divided into three polytomous items: the clock face (worth 2 points), numbers (4 points), and hands (4 points). To develop the items, Rouleau et al. (1992, 1996) completed a "qualitative error analysis" of clocks from 50 participants with dementia (25 with AD, 25 with Huntington's disease), and found six dimensions of errors which they then formalized as items. Rouleau did not report the qualitative methodology for error identification, reliability, or validity of this instrument; it has been found to be sensitive to dementia but less so to MCI (Chiu et al., 2008; Duro et al., 2019).

2.3.2.2. Clock Drawing Interpretation Scale (CDIS)

The Clock Drawing Interpretation Scale (CDIS) is a 20-item rating scale developed from clock drawing errors made by persons with Alzheimer's disease (Mendez et al., 1992). In a sample of forty-six people with dementia recruited from a memory clinic and twenty-six neurologically-healthy older adult controls ($N = 72$), the CDIS was found to have excellent inter-rater reliability ($r = 0.95$), and internal consistency ($r = 0.95$). Its highest concurrent validity was found for constructional praxis, at a moderate level ($r = 0.65-0.66$). The cut score for dementia was 19. Psychometric

properties for individuals with MCI have not been reported for the CDIS.

2.4. Statistical analysis

Descriptive statistics were computed and graphically-assessed for distribution. Internal consistency was assessed using Cronbach's alpha (Cronbach, 1951). Sensitivity and specificity of the two CDT scales to mild impairment were assessed using the receiver operating characteristic area under the curve (ROC AUC) (Zweig and Campbell, 1993) for MoCA cut scores of 26 (Nasreddine et al., 2005) and 23 (Carson et al., 2018). The area under the curve is an estimate of the percentage of times that a true positive is detected (*specificity*) and a false negative is avoided (*sensitivity*), ranging from 0.5 (no better than chance) to 1.0 (accurate 100% of the time). Our two cut scores were chosen to compare the clock's validity against the original mild impairment cut score of 26 (Nasreddine et al., 2005) versus Carson et al.'s (2018) recommendation for better sensitivity in community samples of 23. Finally, we tested the construct validity of the individual items and the three-aspect structure (face, numbers, hands) using Item Response Theory methods to estimate the scales' correspondence to cognitive level. For the Rouleau method's hierarchical polytomous items, we tested the assumptions of dimensionality and ordinal structure using the Graded Response Model (Samejima, 2010), and then tested the amount of information given about cognitive ability using factor analysis with a 3-factor model. For the CDIS's dichotomous items, we tested the information given about cognitive ability using a 2-parameter model (2PL). An alpha level of 0.05 was used for all statistical tests. Analyses were computed in R version 4.2.2 (R Core Team, 2022) using the psych (Revelle, 2022), ltm (Rizopoulos, 2022), and mirt (Chalmers, 2012) packages.

To confirm scoring fidelity, we tested interrater reliability for a subset of cases. The first author selected a random sample of $n = 18$ participants using a simulation-based random number selection in R software, scored these participants' drawings on both scales, then computed weighted kappa statistics for each scale comparing their scores to the original scores in the data set. The kappa statistics were 0.94 for the Rouleau and 0.88 for the CDIS, indicating good to excellent interrater reliability.

3. Results

3.1. Participants

We administered assessments to 345 community-dwelling older adults aged 55 years and older, retaining 310 for this analysis after excluding 35 people with MoCA scores below 20, as individuals with possible dementia are outside this study's focus on mild impairment; this resulted in a final sample of 310 participants. Our sample was predominantly female (75%) and Caucasian (86%), with a mean age 67 (8.9) years, and a mean of 16 (3.2) years of education. The overall mean score on the MoCA was 25 (2.7), on the Rouleau was 9 (1.5), and on the CDIS was 18 (1.9). The sample's MoCA scores were about evenly split between unimpaired/impaired at a cut score of 26; a larger proportion were

classified as unimpaired with a cut score of 23. Table 1 presents demographics and test scores.

3.2. Reliability

Cronbach's alpha (α) was computed to estimate the internal consistency of each clock drawing scale. The Rouleau coefficient α was 0.367 (SD 0.48, 95% CI 0.23, 0.48), indicating poor internal consistency, and the CDIS coefficient α was 0.665 (SD 0.09, 95% CI 0.61, 0.72) indicating moderate internal consistency. Table 2 summarizes these reliability estimates. See Supplementary Tables 1, 2 for item statistics.

3.3. Sensitivity and specificity

We computed ROC curves for both clock drawing scales to establish criterion (or "cut" scores) using the MoCA as the index measure. We computed each scale's area under the curve (AUC), sensitivity (the proportion of scores below the CDT's cut score to our two cut scores on the MoCA) and specificity (the proportion of scores above the cut score on the CDT to the scores above the given cut score). At a MoCA cut of 26, the AUC values were 0.65 for both the Rouleau (0.5945–0.7055) and the CDIS (0.5897–0.7087). At a MoCA cut of 23, these were 0.66 (0.5905–0.7305) for the Rouleau and 0.63 (0.5590–0.6985) for the CDIS. Curves are presented in Figure 1. CDT scale cut scores were determined with a threshold analysis showing the optimal sensitivity and specificity balance. At a MoCA cut of 26, a Rouleau score of 9 (out of 10) had sensitivity of 0.59 and specificity of 0.70, and a CDIS cut score of 19 (out of 20) had sensitivity of 0.78 and specificity of 0.45. At a MoCA cut of 23, a Rouleau score of 9 had sensitivity of 0.66 and specificity of 0.61, and a CDIS cut score of 19 had sensitivity of 0.82 and specificity of 0.37. Table 2 summarizes these statistics.

3.4. Construct validity

3.4.1. Rouleau

Because this scale uses an ordinal item structure, the graded response model (GRM) was used (Samejima, 2010). The GRM estimates the item-wise probabilities of satisfying each criterion (or response level) within each item (progressing from low to high cognitive ability), and a value of the "threshold" between levels, which estimates the item's distinction between each criterion (an important characteristic of an ordinal item). The item information curves (Figures 2A-C) show that the probability of passing more difficult items (y axis) increases with cognitive ability (x axis), but that the level of ability required to pass them is lower than the average cognitive level of our participants. The clearest differentiation between performance on each criterion is in the "hands" item, with distinct differences in the cognitive abilities contained within each score. This suggests that this item clearly discriminates between high, middle, and low cognitive levels, and is therefore a true ordinal item. By contrast, the "face" and "numbers" items' middle scores are nearly contained within the lowest and highest score levels, suggesting that they are less discriminative

TABLE 1 Sample characteristics.

N = 310	MoCA cut of 26			MoCA cut of 23		
	MoCA in unimpaired range (26–30) n = 146 (47%)	MoCA in mild impairment range (20–25) n = 164 (53%)	Test, p-value	MoCA in unimpaired range (23–30) n = 241 (78%)	MoCA in mild impairment range (20–22) n = 69 (22%)	Test, p-value
Sex, n (%)			$\chi^2 = 2.3, NS$			$\chi^2 = 1.9, NS$
Female, 233 (75%)	116 (79%)	117 (71%)		186 (77%)	47 (68%)	
Male, 77 (25%)	30 (21%)	47 (29%)		55 (23%)	22 (32%)	
Race, n (%)			NS*			NS*
White, 266 (86%)	132 (91%)	134 (82%)		217 (90%)	49 (71%)	
Black/African American, 33 (10%)	8 (6%)	25 (15%)		17 (7%)	16 (23%)	
Other, 3 (<1%)	5 (3%)	5 (3%)		6 (2%)	4 (6%)	
Ethnicity, n (%)			NS*			NS*
Hispanic/Latino, 9 (3%)	5 (3%)	4 (2%)		7 (3%)	2 (3%)	
Not Hispanic/Latino, 298 (96%)	140 (97%)	158 (98%)		232 (97%)	66 (97%)	
Age; mean (SD) 69.85 year (8.9)	67 (6.6)	73 (9.8)	$t = -6.03$ <0.001	69 (8.3)	73 (10)	$t = 2.79, 0.006$
Education; mean (SD) 16 year (3.2)	17 (3.2)	15 (2.9)	$t = 4.84, <0.001$	16 (3.2)	14 (2.5)	$t = 6.00, <0.001$
Instrument total scores; mean (SD)**						
MoCA, 25 (2.7)	28 (1.2)	23 (1.6)	$t = 28.51,$ <0.00	26 (2.0)	21 (0.75)	$t = 31.29, <0.00$
Rouleau, 9 (1.5)	9 (1.4)	9 (1.4)	$t = 3.74, <0.001$	9 (1.4)	9 (1.4)	$t = 3.50, <0.00$
CDIS, 18 (1.9)	19 (1.8)	18 (1.9)	$t = 3.44, <0.001$	19 (1.9)	18 (1.8)	$t = 2.26, <0.00$

MoCA, “Montreal Cognitive Assessment”; CDT, “clock drawing test”; CDIS, “Clock Drawing Interpretation Scale”; ADCS, “Alzheimer’s Disease Cooperative Study–Activities of Daily Living Scale.”

* Insufficient group representation to run χ^2 as either full or collapsed factor variable.

** Assessment statistic results are rounded to whole numbers to fit their scoring conventions.

and have very close thresholds—this suggests that, in practice, they are more likely to work as binomial (impaired/unimpaired) than ordinal (step-wise with multiple levels) items. The test information curve (Figure 2D) also shows that this scale yields the most information for participants with lower cognitive performance. The Rouleau items contributed insufficient degrees of freedom to assess goodness-of-fit using root mean square error of approximation (RMSEA) (Cai and Hansen, 2013); the marginal residual were heterogeneous, ranging 5.00 (“face” and “hands”) to 21.74 (“numbers” and “hands”).

To allow comparisons between the Rouleau and CDIS scales, we then ran a factor analysis on the Rouleau scores using a 3-factor item loading model. This model placed the Rouleau scale’s difficulty below the average cognitive level in this sample (−8.28 to −0.54). The “face” aspect was the least discriminating (0.38)

and “numbers” was most discriminating (3.58). Figure 3 shows changes in difficulty across cognitive levels (x axis) for the total score (Figure 3A) and each item (Figure 3B). Note that the curves for “numbers” are shifted to the left, suggesting that this component is easier to score a point on than “hands.” The curve for “face” is virtually flat, suggesting it does not give much discriminating information. The overall leftward shift of the curves shows that this scale provides more information at a lower cognitive level, which is consistent with the graded response modeling above.

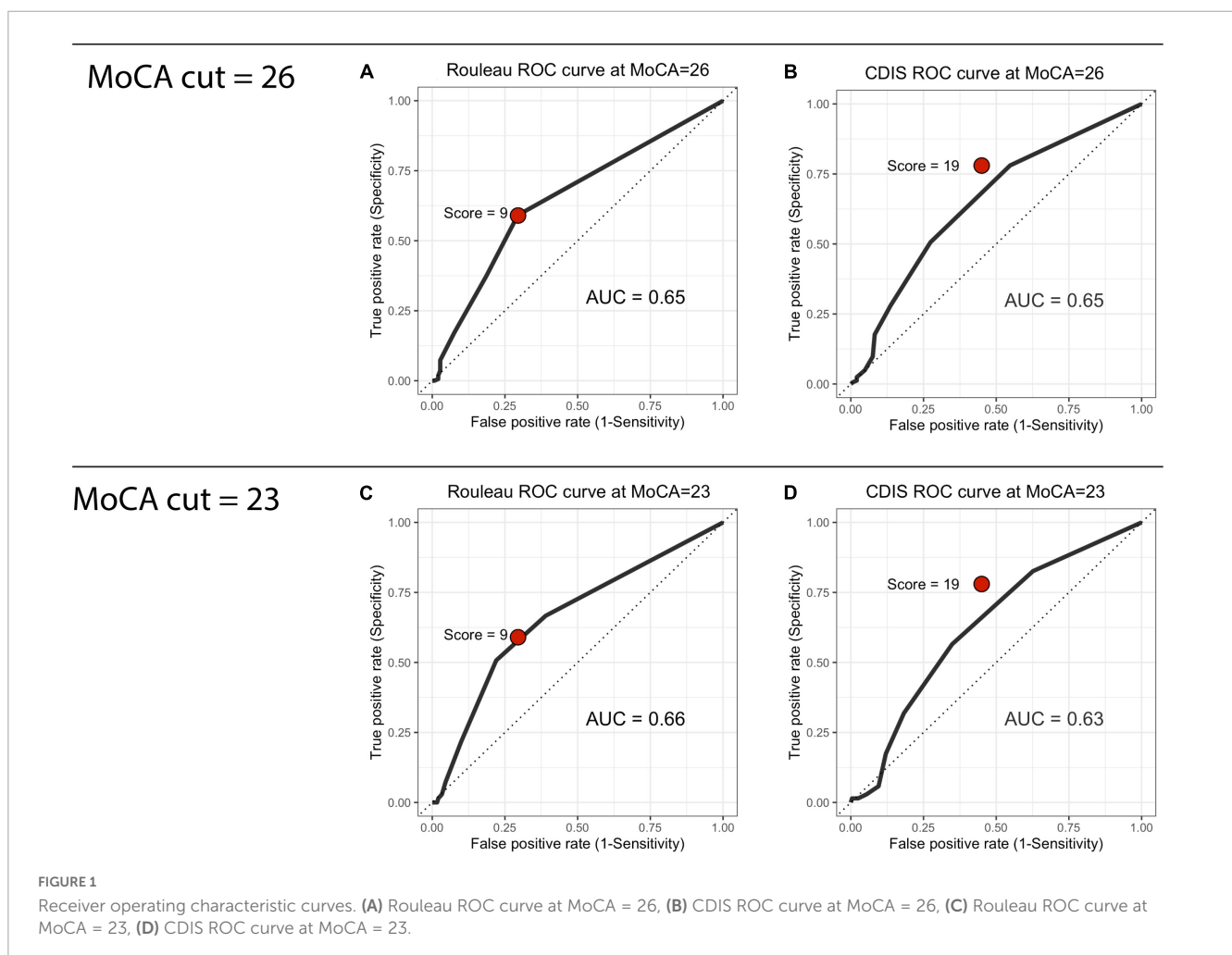
3.4.2. CDIS

This scale’s binomial (0 or 1 point) item structure was assessed using a 2-parameter model (2PL). The 2PL model includes parameter values for item discrimination (α -parameter) and difficulty (β -parameter) to for each CDIS item. The coefficient

TABLE 2 Summary of test statistics.

	MoCA cut score of 26				MoCA cut score of 23			
	Rouleau		CDIS		Rouleau		CDIS	
	Test statistic (sd)	CI	Test statistic (sd)	CI	Test statistic (sd)	CI	Test statistic (sd)	CI
Cronbach's α	0.37 (0.48)	(0.23, 0.48)	0.66 (0.09)	(0.61, 0.72)				
ROC								
AUC	0.68	(0.59, 0.71)	0.65	(0.59, 0.71)	0.66	(0.59, 0.73)	0.63	(0.56, 0.69)
Sensitivity*	0.59		0.78		0.66		0.83	
Specificity*	0.71		0.45		0.61		0.37	

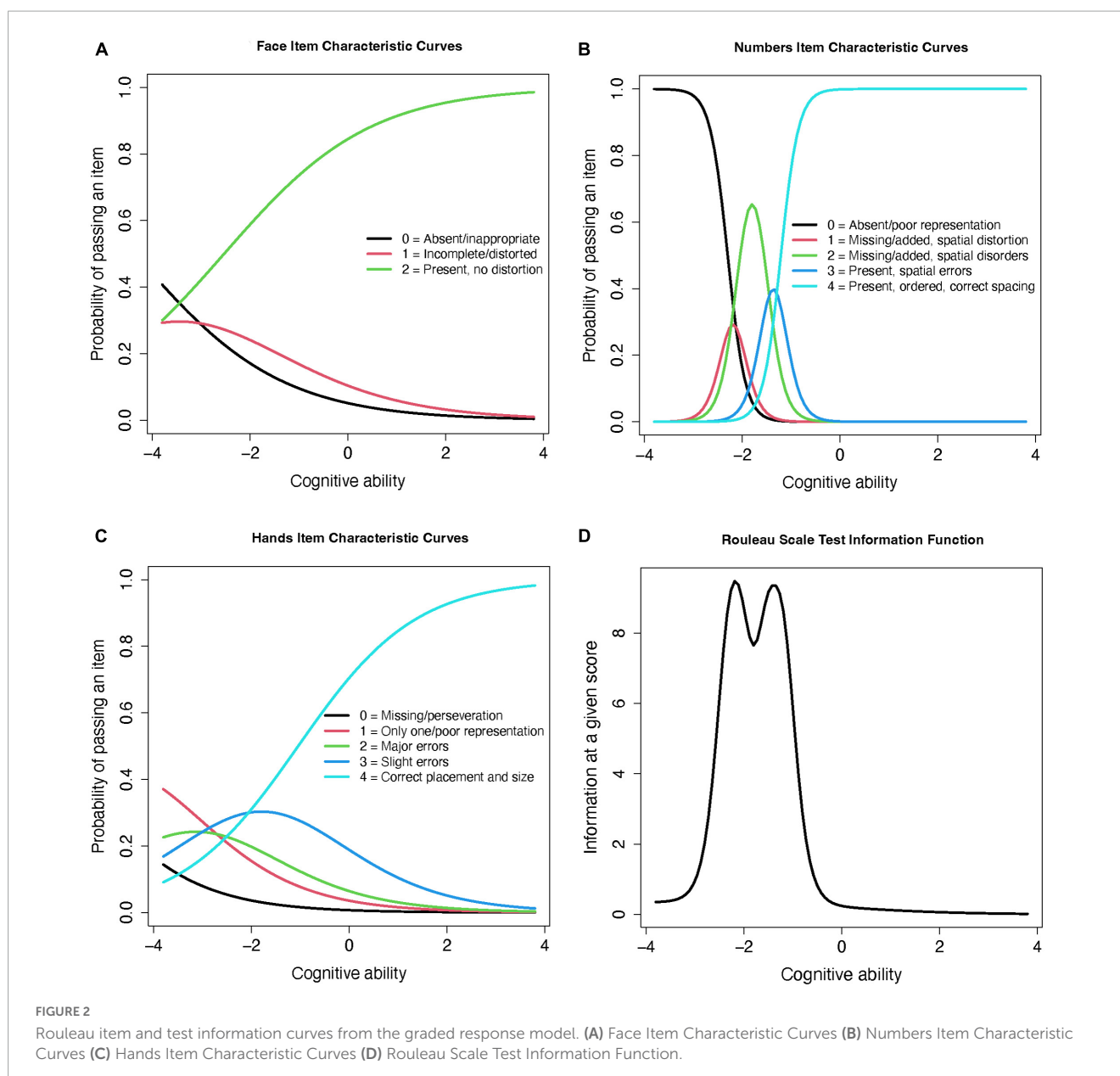
CDIS, "Clock Drawing Interpretation Scale"; ROC, "receiver operating characteristic"; AUC, "Area under the curve." MoCA cut score is not used to compute Cronbach's α .
 * At a score of 9 on the Rouleau and 19 on the CDIS.



values for these estimates are plotted for the items grouped by aspect (Figures 4A-C) and total score (Figure 4D) as a test information curve. As was the case for the Rouleau, this curve shifts to the left; this shows that the CDIS scale yields the most information for participants with poorer cognitive performance, but with more restricted range of abilities compared to the Rouleau scale. Model fit was estimated using RMSEA; a value of 0.09 indicated fair to poor model fit (Boateng et al., 2018).

4. Discussion

In this community-dwelling sample of older adults, we found that the CDIS had moderate reliability (though superior to the Rouleau scale), and that both scoring systems were moderately sensitive to MoCA scores consistent with mild cognitive impairment. Our item response theory models showed that both scales provided very limited information about small



shifts toward impaired cognition, and that the “hands” items appeared to be the most discriminating. These findings are consistent with several prior findings on the Rouleau scale in particular (Chiu et al., 2008; Pinto and Peters, 2009; Talwar et al., 2019). Thus, our hypotheses were partially supported; the CDIS scale was more reliable (but not more sensitive) than the Rouleau scale, and a score of only one unsatisfied criterion had the highest sensitivity to mild impairment. Based on these results, we found insufficient evidence to recommend either scale be used as a stand-alone screening instrument in this population.

Of the two scales, the CDIS showed higher internal consistency, which we expected since reliability tends to increase as the number of items in a scale increases. One possible contribution to our low reliability is item composition. Quantitative ratings of a free-drawn object require some amount of subjective interpretation, so each item must be written for narrow interpretation with unmixed criteria for success (Bandalos, 2018b). Best practices include giving

one criterion for success per item, avoiding qualifiers, and using single clauses (Bandalos, 2018b). Several of these recommendations are violated in these two scales. For instance, the Rouleau method uses qualifiers - such as “most,” “minor,” and “slight”-in 8 out of 10 items, and the CDIS in 3 out of 20. These qualifiers are adjectives expressing relative degree, but neither scale provides metrics to guide clinicians in placing performance between degrees. The influence of wording on reliability could be explored in the future by comparing item-total correlations and item statistics to identify the items with the lowest correlation to the total score, then dropping the low-correlated items and re-analyzing the scales.

Another consideration is our reliability estimator’s underlying assumptions and limitations. We chose Cronbach’s alpha to allow comparisons with prior work (Duro et al., 2019), however, there is debate as to whether this statistic is an appropriate estimator of internal consistency for this type of scale (Cho, 2021; Sijtsma and Pfadt, 2021). First, use of the alpha statistic is based on the

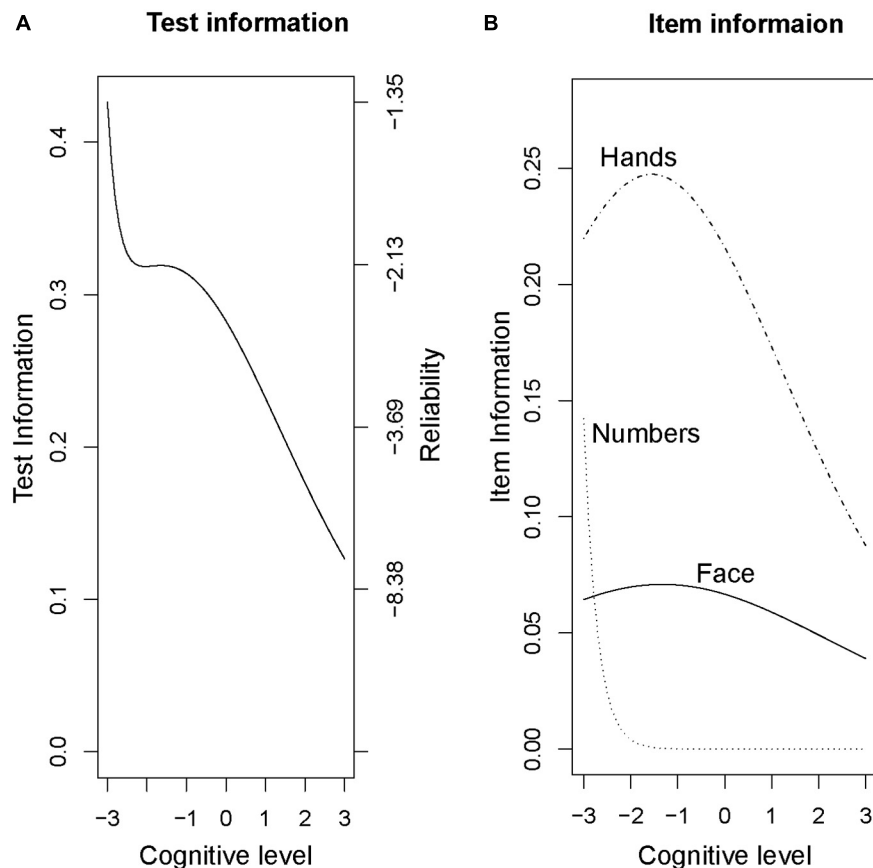


FIGURE 3

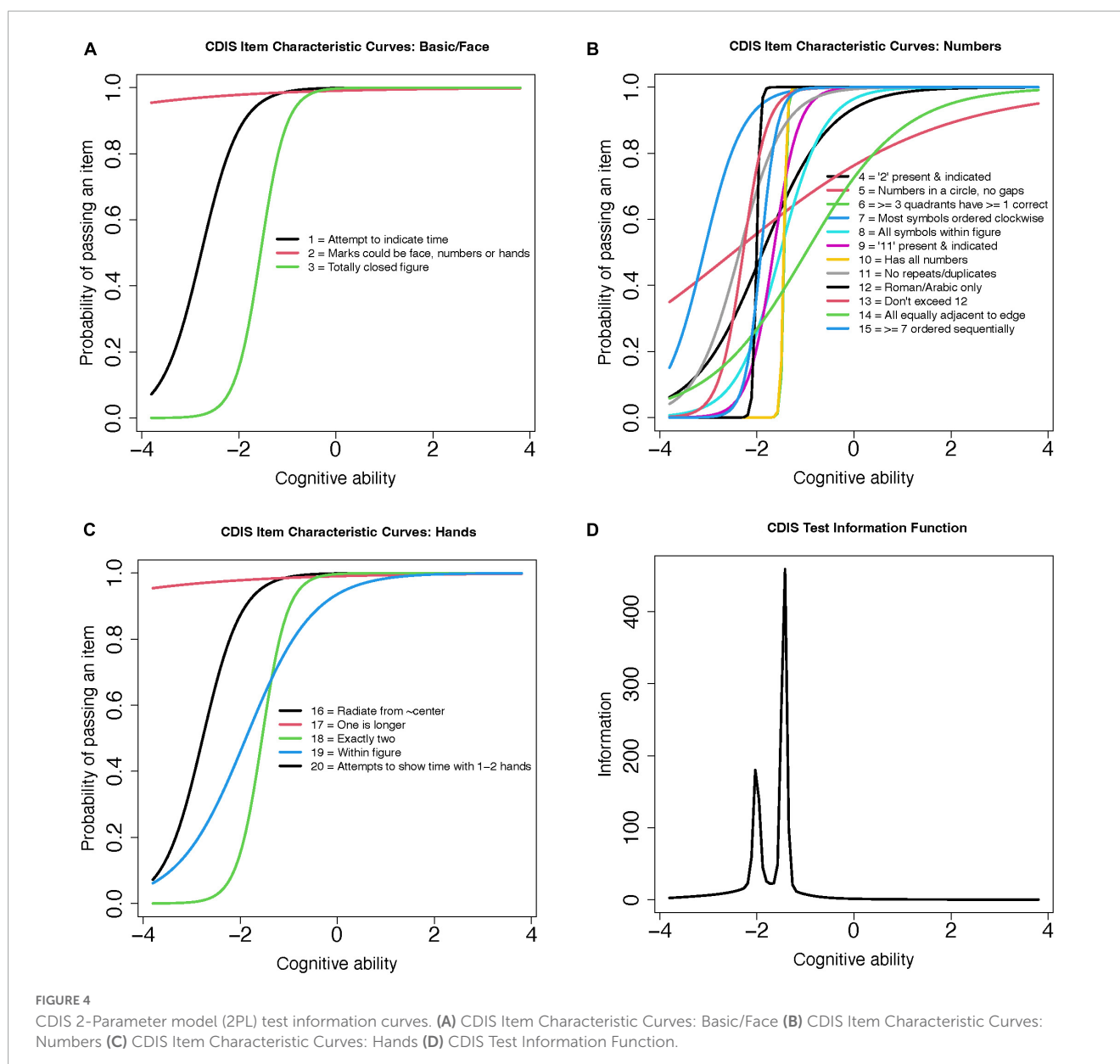
Rouleau factor analysis model, Total and Item Curves. (A) Test information (B) Item information.

assumption that the items of the scale are *unidimensional*, that is, they measure the same construct. Although these scales are screening measures for dementia (one construct), they require many overlapping abilities such as visuospatial, executive function, motor and perceptual integration, and a cultural context for chronological measurement of time (multiple constructs)—thus they may not measure dementia *per se*, but rather multiple skills as proxies of dementia. Second, the *error variance of all items should be uncorrelated*—that is, items should vary independently of one another (Trizano-Hermosilla and Alvarado, 2016). In the assumptions of classical true score theory that underlies alpha, error originates unpredictably from a random process, so item variances should not influence one another. This is unlikely to apply to these scales because elements of a drawing rely on previous elements, for instance the clock's hand placement relies on the clock face being symmetrical and the numbers evenly spaced. Third and finally, the scale's items should be *essentially tau-equivalent*, meaning they may differ in mean and variances but must have similar covariances; if this assumption is violated then alpha is likely to underestimate the scale's true reliability (Bandalos, 2018a). On the other hand, Cronbach's alpha has also been found to mask multidimensionality of test items, and slightly overestimate ordinal scales such as the Rouleau scale (Vaske et al., 2017). We retain Cronbach's alpha for this study as a lower-bound estimate of reliability, and acknowledge that other methods, such as a stratified alpha (estimations for face, numbers, and hands singly),

the greatest lower bound, or expanded factor analytic methods could provide more accurate estimates for this type of scale in a larger sample.

Our criterion validity results were generally consistent with other recent work. At the MoCA cut score of 26, the sensitivity and specificity of these clock scales were moderate, and the AUC point estimates were nearly identical to that of another recent ROC analysis of the Rouleau scale (Duro et al., 2019). Lowering the criterion score to 23 had a negligible effect on AUC, but shifted both clock scales' sensitivity 5–6 points higher, specificity 8–10 points lower, and confidence intervals several points wider, suggesting lower precision. At either cut, the highest sensitivities were lower than the 90% (Nunnally, 1978) to 95% (Kaplan and Sarcuzzo, 2001) recommended for a clinical screening tool, found at a score of only one incorrect item (9/10 on the Rouleau, 19/20 on the CDIS).

There is evidence that the MoCA score of 26 is overly sensitive, resulting in too many false positives (Carson et al., 2018; Townley et al., 2019). With deference to the inconvenience and cost this can cause, we argue that in high-functioning community-dwelling adults it is preferable for a non-invasive and brief screening tool to err on the side of greater sensitivity with the potential for false positives, rather than greater specificity with the potential to overlook cases. The MoCA is a screening measure to detect potential impairment, not to diagnose (Dautzenberg et al., 2020), and administering clinicians will also consider trends, subjective report, and patient concerns in balance with scores. Compared



to the studies recommending lower cut scores (Carson et al., 2018; Milani et al., 2018; Townley et al., 2019), our sample had higher mean education in both impaired and unimpaired groups regardless of classification score, and very low racial and ethnic heterogeneity. Our ROC results at a cut score of 23 may be more valid for community samples with low mean education (Carson et al., 2018) and people from underrepresented racial groups (Milani et al., 2018; Rossetti et al., 2019); however, we could not verify this with our sample.

Our application of item response theory models is the first use of these methods with these two scales. Our results help clarify the relationship between ability and test performance, a key aspect of the test's appropriateness for subtle cognitive decline. In our sample, these scales had levels of difficulty that were more appropriate to rating performance by a person with a lower level of cognition than the early or subtle decline we hope to capture in community-dwelling older adults; in other

words, the scoring criteria on these scales appear to be too easy for this population. We found a ceiling effect that begins to detect change well below an unimpaired level, in turn driving the scales downward in cognitive ability and resulting in restriction of range that masks performances with subtle impairments. We note that both scales' test information curves are bimodal with a leftward shift (right skew), suggesting that they are very informative for a narrow range of ability at a low cognitive level, and not informative for very small decreases from baseline. This is consistent with prior study on their psychometric properties for detecting dementia (Mendez et al., 1992; Chiu et al., 2008) and was also consistent with our and others' ROC analyses (Duro et al., 2019). Hence, many of the items in these scales contributed little to no information toward characterizing subtle change in our sample. A caveat is that the GRM and 2PL models had poor indicators of fit—the model “expects” a response pattern that differed from that of our sample. We suspect two sources of

this high error approximation; the scales may contain poorly-performing items that introduce noise into the model, and, the task's step-by-step nature may constitute multidimensional latent constructs contributing to each item. Both of these hypotheses could be explored using exploratory factor analysis or other latent trait modeling in a future study.

It is also notable to consider the change in diagnostic definition and practice since these scales were developed. These scales' purpose was to detect impairment within a binomial "dementia vs. no dementia" construct of cognitive function, which contrasts with our current understanding of neurocognitive decline as a continuum. For instance, the original validation paper for the CDIS compared healthy older adults to persons with dementia but excluded those with "questionable" cognitive impairment, establishing just one cut score for dementia rather than a range of scores across the cognitive spectrum. Similarly, the scoring criteria for the Rouleau scale (Rouleau et al., 1992, 1996) were developed according to errors made by people with dementia, and did not characterize a range of unimpaired performance as a reference. At that time, a range of cognitive function was not necessarily considered clinically useful, but this is no longer the case.

Based on our findings, we suggest that the task of drawing a clock is not "too easy" for a person with subtle cognitive deficits to execute without errors, but rather that dementia-validated paper scales used to score their execution may be too lenient. If the scales' sensitivity and specificity to MCI could be improved, they would meet established criteria for good clinical screening tools (Edwards et al., 2019). Future studies could attempt to improve these scales by repeating testing and analysis after removing items with low item-total correlations and RMSEA, conducting expanded factor analyses, and using alternative reliability estimators.

4.1. Limitations

Our study has several limitations. First, our sample is not representative of the population of normally-aging community-dwelling older adults. We had three female participants for every male, and an insufficient number of participants who were Black, Hispanic, or other racial or ethnic identities to compute statistics on these sub-groups. Although our racial and ethnic demographic imbalance isn't a barrier to comparing these results with the original validation studies for these scales (neither reported these variables), it is a barrier to implementing these findings in broader clinical practice because it does not reflect the proportions of these individuals in the population of people who develop dementia, nor does it inform us on bias in these instruments (Chin et al., 2011; Weuve et al., 2018) particularly given our index measure's different score distributions in some populations (O'Caomh et al., 2016; Milani et al., 2018; Rossetti et al., 2019). A second limitation is the lack of cognitive diagnoses for our sample. We used the MoCA (a screening measure with a disputed cut score) to test the sensitivity of the two CDT scales (another screening measure with no established cut score) without scores from a diagnostic criterion measure. Thus, our analyses are limited to observing groups with scores *consistent with*, but not definitively diagnostic of, unimpaired versus mildly impaired cognition. Finally, our design precludes determination of predictive validity because we used a cross-sectional design without a repeating measure.

5. Conclusion

Although the task of free-drawing a clock has been found to represent cognitive abilities affected by MCI (Royall et al., 1999; Yuan et al., 2021) our analysis of the CDIS and Rouleau clock drawing scales found that they did not reliably identify persons scoring in the mildly impaired range on the MoCA. We recommend use of these scales in combination with other screening tools primarily as an opportunity to observe visuospatial and executive abilities, with awareness of a ceiling effect in the interpretation of their numeric score.

Cognitive changes are often subtle and granular; even if diagnosis is categorical, small changes are both clinically meaningful for care-planning and functionally meaningful for patients. Assessment tools must reflect small differences in function to be useful for screening community-dwelling older adults. In our community-dwelling sample, the Rouleau and CDIS clock drawing scales did not reflect these small differences.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the University of Wisconsin-Madison Health Sciences Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

KK-F, TM, GG, and DE: concept, data collection, and design. KK-F: statistical analysis and writing. TM, GG, and DE: editing. All authors contributed to the manuscript revision, and approved the final submission.

Funding

This work was supported by a TL1 trainee grant to the first author from 2021 (5TL1TR002375-05) to 2023 (2TL1TR002375-06), the GG Fund 233 PRJ22AW 176000, and the Marsh Center Fund 233 PRJ56ZI 176000 from the University of Wisconsin-Madison Occupational Therapy Program and Kinesiology Department.

Acknowledgments

We authors are grateful to the participants who shared their time for this study, and for previous students in the Functional Cognition Lab. Program support was provided by the University

of Wisconsin-Madison, School of Education, Department of Occupational Therapy. The first author extends gratitude for mentorship from Daniel Bolt in the University of Wisconsin-Madison, School of Education, Dept. Education Psychology, and for training and program support from the University of Wisconsin-Madison, School of Medicine and Public Health, Institute for Clinical and Translational Sciences. We would like to thank the Marsh Center for Exercise and Movement Research at the University of Wisconsin-Madison for their generous support for this work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- American Psychiatric Association [APA] (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. Arlington, VA: American Psychiatric Association.
- Anderson, N. D. (2019). State of the science on mild cognitive impairment (MCI). *CNS Spectr.* 24, 78–87.
- Arevalo-Rodriguez, I., Smailagic, N., Roqué-Figuls, M., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., et al. (2021). Mini-Mental State Examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). *Cochr. Datab. Syst. Rev.* 7:Cd010783.
- Bandalos, D. L. (2018a). "Introduction to reliability and the classical test theory model," in *Measurement theory and applications for the social sciences*, (New York, NY: Guilford Press), 155–210.
- Bandalos, D. L. (ed.). (2018b). *Measurement theory and applications for the social sciences*. New York, NY: Guilford Press.
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., and Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Front. Public Health* 6:149. doi: 10.3389/fpubh.2018.00149
- Cai, L., and Hansen, M. (2013). Limited-information goodness-of-fit testing of hierarchical item factor models. *Br. J. Math. Stat. Psychol.* 66, 245–276. doi: 10.1111/j.2044-8317.2012.02050.x
- Carson, N., Leach, L., and Murphy, K. J. (2018). A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *Int. J. Geriatr. Psychiatry* 33, 379–388. doi: 10.1002/gps.4756
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *J. Stat. Softw.* 48, 1–29.
- Chin, A. L., Negash, S., and Hamilton, R. (2011). Diversity and disparity in dementia: The impact of ethnic differences in Alzheimer disease. *Alzheimer Dis. Assoc. Disord.* 25, 187–195.
- Chiu, Y.-C., Li, C.-L., Lin, K.-N., Chiu, Y.-F., and Liu, H.-C. (2008). Sensitivity and specificity of the clock drawing test, incorporating Rouleau scoring system, as a screening instrument for questionable and mild dementia: Scale development. *Int. J. Nurs. Stud.* 45, 75–84. doi: 10.1016/j.ijnurstu.2006.09.005
- Cho, E. (2021). Neither Cronbach's alpha nor McDonald's omega: A commentary on Sijtsma and Pfadt. *Psychometrika* 86, 877–886. doi: 10.1007/s11336-021-09801-1
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334.
- Dautzenberg, G., Lijmer, J., and Beekman, A. (2020). Diagnostic accuracy of the Montreal Cognitive Assessment (MoCA) for cognitive screening in old age psychiatry: Determining cutoff scores in clinical practice. Avoiding spectrum bias caused by healthy controls. *Int. J. Geriatr. Psychiatry* 35, 261–269. doi: 10.1002/gps.5227
- Davoudi, A., Dion, C., Amini, S., Libon, D. J., Tighe, P. J., Price, C. C., et al. (2020). Phenotyping cognitive impairment using graphomotor and latency features in digital clock drawing test. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2020, 5657–5660. doi: 10.1109/EMBC44109.2020.9176469
- Dion, C., Arias, F., Amini, S., Davis, R., Penney, D., Libon, D. J., et al. (2020). Cognitive correlates of digital clock drawing metrics in older adults with and without mild cognitive impairment. *J. Alzheimers Dis.* 75, 73–83.
- Dion, C., Tanner, J. J., Formanski, E. M., Davoudi, A., Rodriguez, K., Wiggins, M. E., et al. (2022). The functional connectivity and neuropsychology underlying mental planning operations: Data from the digital clock drawing test. *Front. Aging Neurosci.* 14:868500. doi: 10.3389/fnagi.2022.868500
- Duro, D., Freitas, S., Tábuas-Pereira, M., Santiago, B., Botelho, M. A., and Santana, I. (2019). Discriminative capacity and construct validity of the clock drawing test in mild cognitive impairment and Alzheimer's disease. *Clin. Neuropsychol.* 33, 1159–1174. doi: 10.1080/13854046.2018.1532022
- Edwards, D. F., Al-Heizan, M. O., and Giles, G. M. (2019). "Ch. 6: Baseline cognitive screening tools," in *Functional cognition and occupational therapy: A practical approach to treating individuals with cognitive loss*, eds T. Wolf, D. Edwards, and G. Giles (Bethesda, MD: AOTA Press).
- Esteban-Santillan, C., Praditsuwan, R., Veda, H., and Geldmacher, D. S. (1998). Clock drawing test in very mild Alzheimer's disease. *J. Am. Geriatr. Soc.* 46, 1266–1269.
- Hazan, E., Frankenburg, F., Brenkel, M., and Shulman, K. (2018). The test of time: A history of clock drawing. *Int. J. Geriatr. Psychiatry* 33, e22–e30. doi: 10.1002/gps.4731
- Jekel, K., Damian, M., Wattmo, C., Hausner, L., Bullock, R., Connelly, P. J., et al. (2015). Mild cognitive impairment and deficits in instrumental activities of daily living: A systematic review. *Alzheimers Res. Ther.* 7:17.
- Kaplan, R. M., and Sarcuzzo, D. P. (2001). *Psychological testing: Principles, applications, and issues*. Belmont, CA: Wadsworth.
- Lam, A. Y., Anderson, K., Borson, S., and Smith, F. L. (2011). A pilot study to assess cognition and pillbox fill accuracy by community-dwelling older adults. *Consult Pharm.* 26, 256–263. doi: 10.4140/TCP.n.2011.256
- Lin, J. S., O'Connor, E., Rossom, R. C., Perdue, L. A., Burda, B. U., Thompson, M., et al. (2013). *U.S. Preventive Services task force evidence syntheses, formerly systematic evidence reviews. Screening for cognitive impairment in older adults: An evidence update for the US preventive services task force*. Rockville, MD: Agency for Healthcare Research and Quality (US).
- Lindbergh, C. A., Dishman, R. K., and Miller, L. S. (2016). Functional disability in mild cognitive impairment: A systematic review and meta-analysis. *Neuropsychol. Rev.* 26, 129–159.
- Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., et al. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* 396, 413–446.
- Lowe, D. A., and Linck, J. F. (2021). Item response theory analysis of the texas functional living scale. *Arch. Clin. Neuropsychol.* 36, 135–144.
- Luo, H., Andersson, B., Tang, J. Y. M., and Wong, G. H. Y. (2020). Applying item response theory analysis to the montreal cognitive assessment in a low-education older population. *Assessment* 27, 1416–1428. doi: 10.1177/1073191118821733
- Mendez, M. F., Ala, T., and Underwood, K. L. (1992). Development of scoring criteria for the clock drawing task in Alzheimer's disease. *J. Am. Geriatr. Soc.* 40, 1095–1099.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnagi.2023.1210585/full#supplementary-material>

- Milani, S. A., Marsiske, M., Cottler, L. B., Chen, X., and Striley, C. W. (2018). Optimal cutoffs for the Montreal Cognitive Assessment vary by race and ethnicity. *Alzheimers Dement.* 10, 773–781. doi: 10.1016/j.dadm.2018.09.003
- Müller, S., Preische, O., Heymann, P., Elbing, U., and Laske, C. (2017). Increased diagnostic accuracy of digital vs conventional clock drawing test for discrimination of patients in the early course of Alzheimer's disease from cognitively healthy individuals. *Front. Aging Neurosci.* 9:101. doi: 10.3389/fnagi.2017.00101
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x
- Nunnally, J. C. (1978). *Psychometric theory*, 2nd Edn. New York, NY: McGraw-Hill.
- O'Caomh, R., Timmons, S., and Molloy, D. W. (2016). Screening for Mild Cognitive Impairment: Comparison of "MCI Specific" Screening Instruments. *J. Alzheimers Dis.* 51, 619–629.
- Petersen, R. C., Lopez, O., Armstrong, M. J., Getchius, T. S. D., Ganguli, M., Gloss, D., et al. (2018). Practice guideline update summary: Mild cognitive impairment: Report of the guideline development, dissemination, and implementation subcommittee of the American academy of neurology. *Neurology* 90, 126–135.
- Pinto, E., and Peters, R. (2009). Literature review of the clock drawing test as a tool for cognitive screening. *Dement. Geriatr. Cogn. Disord.* 27, 201–213.
- R Core Team (2022). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Revelle, W. (2022). *psych: Procedures for psychological, psychometric, and personality research. Version 2.2.5*.
- Rizopoulos, D. (2022). *ltm: Latent Trait Models under IRT. Version 1.2-0*.
- Rossetti, H. C., Smith, E. E., Hynan, L. S., Lacritz, L. H., Cullum, C. M., Van Wright, A., et al. (2019). Detection of mild cognitive impairment among community-dwelling african americans using the Montreal Cognitive Assessment. *Arch. Clin. Neuropsychol.* 34, 809–813. doi: 10.1093/arclin/acy091
- Rouleau, I., Salmon, D. P., and Butters, N. (1996). Longitudinal analysis of clock drawing in Alzheimer's disease patients. *Brain Cogn.* 31, 17–34.
- Rouleau, I., Salmon, D. P., Butters, N., Kennedy, C., and McGuire, K. (1992). Quantitative and qualitative analyses of clock drawings in Alzheimer's and Huntington's disease. *Brain Cogn.* 18, 70–87.
- Royall, D. R., Mulroy, A. R., Chiodo, L. K., and Polk, M. J. (1999). Clock drawing is sensitive to executive control: A comparison of six methods. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 54B, 328–333. doi: 10.1093/geronb/54b.5.p328
- Samejima, F. (2010). "The general graded response model," in *Handbook of polytomous item response theory models*, eds M. Nering and R. Ostini (New York, NY: Routledge/Taylor and Francis Group), 77–107.
- Shulman, K. I. (2000). Clock-drawing: Is it the ideal cognitive screening test? *Int. J. Geriatr. Psychiatry* 15, 548–561.
- Sijtsma, K., and Pfadt, J. M. (2021). Part II: On the use, the misuse, and the very limited usefulness of Cronbach's alpha: Discussing lower bounds and correlated errors. *Psychometrika* 86, 843–860. doi: 10.1007/s11336-021-09789-8
- Strauss, E., Sherman, E. M. S., and Spreen, O. (eds) (2006). "Tests of visual perception; clock drawing test," in *A compendium of neuropsychological tests: Administration, norms, and commentary*, 3rd Edn (New York, NY: Oxford University Press), 972–982.
- Sunderland, T., Hill, J. L., Mellow, A. M., Lawlor, B. A., Gundersheimer, J., Newhouse, P. A., et al. (1989). Clock drawing in Alzheimer's disease. *J. Am. Geriatr. Soc.* 37, 725–729.
- Talwar, N. A., Churchill, N. W., Hird, M. A., Pshonyak, I., Tam, F., Fischer, C. E., et al. (2019). The neural correlates of the clock-drawing test in healthy aging. *Front. Hum. Neurosci.* 13:25. doi: 10.3389/fnhum.2019.00025
- Townley, R. A., Syrjanen, J. A., Botha, H., Kremers, W. K., Aakre, J. A., Fields, J. A., et al. (2019). Comparison of the short test of mental status and the Montreal Cognitive Assessment across the cognitive spectrum. *Mayo Clin. Proc.* 94, 1516–1523.
- Trizano-Hermosilla, I., and Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Front. Psychol.* 7:769. doi: 10.3389/fpsyg.2016.00769
- Tsolaki, M., Kounti, F., Agogiatou, C., Poptsi, E., Bakoglidou, E., Zafeiropoulou, M., et al. (2011). Effectiveness of nonpharmacological approaches in patients with mild cognitive impairment. *Neurodegener. Dis.* 8, 138–145.
- Vaske, J., Beaman, J. J., and Sponarski, C. C. (2017). Rethinking internal consistency in Cronbach's alpha. *Leisure Sci.* 39, 163–173.
- Weuve, J., Barnes, L. L., Mendes de Leon, C. F., Rajan, K. B., Beck, T., Aggarwal, N. T., et al. (2018). Cognitive aging in black and white Americans: Cognition, cognitive decline, and incidence of Alzheimer disease dementia. *Epidemiology* 29, 151–159.
- Yuan, J., Libon, D. J., Karjadi, C., Ang, A. F. A., Devine, S., Auerbach, S. H., et al. (2021). Association between the digital clock drawing test and neuropsychological test performance: Large community-based prospective cohort (Framingham Heart Study). *J. Med. Int. Res.* 23:e27407. doi: 10.2196/27407
- Zweig, M. H., and Campbell, G. (1993). Receiver-Operating Characteristic (ROC) plots - a fundamental evaluation tool in clinical medicine. *Clin. Chem.* 39, 561–577.

Appendix

Appendix A Rouleau and CDIS scale score rubrics.

Rouleau system.

Participant Code		
1. Integrity of the clock face (maximum; 2 points)		
	2: Present without gross distortion	
	1: Incomplete or some distortion	
	0: Absent or totally inappropriate	
2. Presence and sequencing of the numbers (maximum: 4 points)		
	4: All present in the right order and at most minimal error in the spatial arrangement	
	3: All present but errors in spatial arrangements	
	2: Numbers missing or added but no gross distortions of the remaining numbers; numbers placed in clockwise direction; numbers all present, but gross disorders of spatial layout (i.e., hemineglect, numbers outside the clock)	
	1: Missing or added numbers and gross spatial distortion	
	0: Absence or poor representation of numbers	
3. Presence and placement of hands (maximum: 4 points)		
	4: Hands are in the correct position and the size difference is respected	
	3: Slight errors in the placement of the hands or no representation of size difference between the hands	
	2: Major errors in the placement of the hands (significantly out of course including 10 to 11)	
	1: Only one hand or poor representation of two hands	
	0: No hands or perseveration on hands	
	Total	

Clock Drawing Interpretation Scale (CDIS).

Participant Code		
	Criteria	Point
1	There is an attempt to indicate a time in any way	
2	All marks or items can be classified as either part of a closure figure, a hand, or a symbol for clock numbers.	
3	There is a totally closed figure without gaps (closure figure)	
Score only if symbols for clock numbers arc present		
4	A "2" is present and is pointed out <i>in</i> some way for the time	
5	Most symbols are distributed as a circle without major saps.	
6	Three or more clock quadrants have one or more appropriate numbers: 12–3, 3–6, 6–9, 9–12 per respective clockwise quadrant.	
7	Most symbols are ordered in a clockwise or rightward direction.	
8	All symbols are totally within a closure figure	
9	An "11" is present and is pointed out in some way for the time.	
10	All numbers 1–12 are indicated.	
11	There are no repeated or duplicated symbols	
12	There are no substitutions for Arabic or Roman numerals.	
13	The numbers do not go beyond the number 12	
14	All symbols lie about equally adjacent to a closure figure edge,	
15	Seven or more of the same symbol type are ordered sequentially.	
Score Only if One or More Hands Are Present		
16	All hands radiate from the direction of a closure figure center.	
17	One hand is visibly longer than another hand	
18	There are exactly two distinct and separable hands.	
19	All hands are totally within a closure figure.	
20	There is an attempt to indicate a time with one or more hands.	
	Total (out of 20)	