

Clock Drawing in the Montreal Cognitive Assessment: Recommendations for Dementia Assessment

Catherine C. Price^{a, b, d, e} Holly Cunningham^a Nicole Coronado^a
Alana Freedland^a Stephanie Cosentino^{g-i} Dana L. Penney^{j, k} Alfio Penisi^l
Dawn Bowers^{a, c-f} Michael S. Okun^{c-f} David J. Libon^m

Departments of ^aClinical and Health Psychology, ^bAnesthesiology, and ^cNeurology, University of Florida, ^dCenter for Movement Disorders and Neurorestoration, University of Florida, ^eNational Parkinson Foundation Center of Excellence, University of Florida, and ^fMcKnight Brain Institute, University of Florida, Gainesville, Fla., ^gGertrude H. Sergievsky Center, ^hTaub Institute for Research on Alzheimer's Disease and the Aging Brain, and ⁱDepartment of Neurology, Columbia University Medical Center, New York, N.Y., ^jDepartment of Neurology, Lahey Clinic, Burlington, Mass., ^kDepartment of Neurology, Tufts University School of Medicine, Boston, Mass., ^lSouth Jersey Radiology Associates, Voorhees, N.J., and ^mDepartment of Neurology, Drexel University College of Medicine, Philadelphia, Pa., USA

Key Words

Montreal Cognitive Assessment scoring, rater reliability · Parkinson disease with dementia · Alzheimer disease · Vascular dementia · Cognition

scoring with MoCA criteria requires practice. Supplementing a clock copy to the standard MoCA test (takes <1 min) will improve dementia assessment.

Copyright © 2011 S. Karger AG, Basel

Abstract

Background: Clock drawing is part of the Montreal Cognitive Assessment (MoCA) test but may have administration and scoring limitations. We assessed (1) the reliability of the MoCA clock criteria relative to a published error scoring approach, (2) whether command-only administration could distinguish dementia from cognitively intact individuals and (3) the value of adding a clock copy condition to the MoCA.

Methods: Three novice raters and clocks from dementia and control participants were used to assess the 3 aims. **Results:** MoCA interrater and intrarater reliability were low (i.e. intra-class correlation coefficient = 0.12–0.31) and required repeat training. Clocks drawn to command classified dementia at chance. Inclusion of a copy condition demonstrated expected dementia subgroup patterns. **Conclusion:** Reliable clock

Introduction

Clock drawing has historical ties to neurology [1, 2] and assesses diverse cognitive functions in addition to visuoconstruction [2–6]. Drawing a clock from memory and setting the hands to a specific time requires at least auditory comprehension, the ability to persist in drawing, remembering task instructions, and the ability to translate visuo-perceptual information into an effective motor act (e.g. construction). These cognitive functions are often compromised with dementia. Consequently, clock drawing has gained wide acceptance as a useful tool for evaluating dementia and dementia subtypes [4, 7].

Clock drawing is considered particularly beneficial for dementia assessment when both command and copy

conditions are used and analyzed for errors [5, 8, 9]. As classically defined by Edith Kaplan (1988), clock test administration involves first *commanding* patients to ‘draw the face of a clock with all the numbers and set the two hands to 10 after 11’. On a separate sheet, patients then *copy* a predrawn clock model. While the command condition requires numerous cognitive functions, the copy condition largely draws upon visuospatial and executive functions [5, 10]. Individuals with deficits in visuospatial and executive functions typically perform poorly on command and copy conditions [10]. They present with an inability to improve in their drawing despite the presentation of a clock model; errors from the command condition are transferred to the copy condition [5, 8, 10]. In dementia, poor command and copy performance is often seen among individuals with frontal system deficits like Parkinson’s disease with dementia (PDD) or Huntington’s disease [5, 10, 11]. By contrast, individuals with less executive dysfunction relative to other domains of cognitive impairment (e.g. Alzheimer’s disease, AD) typically improve from command to copy [3, 5, 9, 10, 12, 13]. Thus, when command and copy conditions are used together, clock drawing appears particularly helpful for differentiating dementia subtypes.

Presently, clock drawing continues to be used at the bedside and in the community for dementia screening. Within medical settings, general practitioners and medical residents with little training in cognitive principles or test psychometry often employ clock drawing tests with other cognitive screeners. Based on our observations of novice clinicians using clock drawing, we identified concerns with scoring and administration. These concerns stimulated the present investigation.

First, clinicians using clock drawing rarely consider rater/scorer reliability. This should be a primary consideration when employing cognitive tests. Without assessing one’s reliability for test scoring, there is questionable validity. This makes it difficult to accurately follow patients longitudinally for clinical change or to compare patient groups. Consider the modern dementia screening test, the Montreal Cognitive Assessment (MoCA [14]), which has gained popularity due to its clinical value over the Mini Mental State Examination (MMSE) [15, 16]. Within the MoCA, clock drawing is one test item involving 3 of the total 30 points possible. Likely to maximize clinical time, clock scoring criteria are basic with 1 point per clock contour, numbers and hands. Criteria appear open to interpretation, e.g. ‘the clock face must be a circle with only minor distortion acceptable (e.g. slight imperfection on closing the circle)’. Thus, novice raters may

have difficulty reliably scoring clocks. In stark contrast, clock scoring within the field of neuropsychology employs an analysis of errors, for error type (e.g. semantic, graphomotor, perseverative) informs clinicians about lesions/pathology [3, 10, 11, 17]. Cosentino et al. [10] provide one error analysis approach based on cognitive theory and common dementia pathology. Templates and precise error definitions are used in order to reduce rater error. Like the MoCA clock drawing, however, novice clinician reliability for this in-depth error analysis has yet to be investigated. For these reasons, our first study aim investigated novice rater intra- and interrater reliability for the MoCA clock scoring paradigm and the Cosentino et al. error analysis approach [10]. Due to the abbreviated MoCA scoring guidelines, we hypothesized that raters would find it more difficult to achieve adequate rater reliability with the MoCA scoring paradigm relative to the error scoring paradigm.

Second, clock drawing administration is frequently truncated to the command condition. We find this surprising given that the copy condition takes only approximately 1 min to administer and has literature attesting to its useful clinical properties. Even the MoCA [14] truncates the item of clock drawing to the command condition. Although the MoCA was designed to help identify dementia and not diagnose subtypes, supplementing a clock copy to the MoCA test may promote better understanding of the patient’s underlying disorder. Our second study aim examined whether the command-only condition was sufficient for discriminating dementia patients from cognitively intact adults. We examined this question with both the MoCA scoring paradigm and the more in-depth error analysis identified by Cosentino et al. [10]. For both scoring paradigms, we hypothesized that the total command condition score alone would not sufficiently differentiate dementia from cognitively intact adults.

Third, there has been no investigation as to whether classic neuropsychological in-depth error scoring paradigms are required to reveal command to copy patterns seen in subcortical or frontal system dementias relative to other dementia forms such as AD. Given the robust literature on command versus copy cognitive demands, we hypothesized that MoCA scored clocks would demonstrate command-copy differences in dementia subtypes.

The overall goal of the current investigation was to improve the clinical utility of clock drawing in clinical settings for dementia assessment. We examined the need to establish rater reliability, the usefulness of the clock

command condition for differentiating dementia from controls, and whether error-based scoring is necessary to identify command-copy patterns in AD versus frontal-subcortical dementia subgroups.

Methods

Participants and Methods

This study was approved by the University of Florida Institutional Review Board and followed Declaration of Helsinki principles. The investigation used data already collected from clinical research investigations or from patients seen clinically as part of a university-based memory disorder clinic. There were 3 aims for the investigation. Aim 1 assessed rater reliability for clock scoring techniques. Aims 2 and 3 specifically examined the value of clock command and copy conditions in dementia and controls.

Aim 1

The first aim examined rater reliability for two different clock scoring methods. The primary investigator randomly selected 160 clock data from the files of 355 individuals diagnosed with idiopathic nondemented Parkinson's disease [18], probable PDD [19], AD [20], dementia associated with small-vessel vascular disease but no major vessel strokes (VaD) [21], or no dementia. These 160 clocks were then randomly organized into 4 separate sets (A–D; 40 clocks/set). Three novices to clock scoring and naïve to hypotheses then studied scoring criteria (MoCA guidelines, see www.mocatest.org; Cosentino criteria [10], see clock scoring criteria below).

Interrater Reliability

The interrater reliability was examined by having each rater independently score the 4 clock sets (320 clocks/rater; 960 clocks total; scored in order A–D, one set completed before the next one initiated). Following reliability analysis, discrepancies were resolved between the raters before moving on to the next set.

Intrarater Reliability

The intrarater reliability was assessed by having raters independently score the first clock set (set A) on 2 separate occasions using both scoring methods. If reliability was not achieved with the first set, the rater measured a second set, etc.

Aim 2

The second aim examined whether the clock criteria scores for the command condition alone could distinguish dementia patients from nondemented controls, as well as AD from frontal-subcortical dementia (PDD, VaD) subgroups. This aim used new clock sets from 231 individuals who had been diagnosed clinically with dementia as well as 50 nondemented controls. The clock data were acquired from clinical charts (nonconsecutive manner) and from nondemented controls enrolled in research investigations with thorough cognitive screening. To ensure the accuracy of dementia versus control classification, this aim only used clock data from individuals who had been diagnosed with dementia via a memory disorder center consensus panel (behavioral neurolo-

gist, neurologist with specialty in movement disorders, neuropsychologist, geriatrician, social worker) and had met DSM-IV dementia criteria [22]. By contrast, the nondemented controls were required to be cognitively intact. They had to have little to no medical comorbidity as measured by the Charlson Comorbidity Index (score = 0–33; higher = more comorbidity [23]), achieve at least an average score on a standardized intelligence test (i.e. 100 ± 15 on the Wechsler Abbreviated Intelligence Scale [24]), and perform within the average range on a set of neuropsychological tests assessing memory, language, category fluency, visuospatial, processing speed and inhibitory functions. Clock drawing had not been used to diagnose or classify clinical groups. A rater with high reliability (rater No. 2; see results for aim 1 reliability scores) completed all MoCA clock measurements. Two raters with high intra- and interrater reliability (raters No. 1 and 2; see aim 1 results) completed the Cosentino criteria measurements. Raters were blind to diagnoses.

To examine whether clock criteria scores for the command condition alone could classify AD from subcortical dementias, the dementia group was further divided into individuals who had met criteria for AD [20] ($n = 73$), VaD [21] ($n = 25$) and PDD [19] ($n = 18$). Due to the increased knowledge of mixed pathology in AD and VaD [25, 26], the AD and VaD patients were required to have had a brain MRI that could then be rated for presence of white-matter disease/white-matter hyperintensities by a trained neuroradiologist (A.P.). Based on ratings from the Junque Visual Rating Scale (rater reliability >0.99 ; severity range from 0 to 40, 40 = maximum [27]) and previous research using this scale [28–30], we a priori required that participants with AD have minimal evidence of white-matter changes suggestive of small-vessel vascular pathology on MR images (Junque score <8 ; final $n = 73$), while VaD patients had to meet criteria for VaD and evidence of severe white-matter changes (i.e. Junque rating scale >18 ; final $n = 25$ [28]). It was also required that the Ischemic Scale scores for these patient groups were below the threshold often associated with a diagnosis of multi-infarct dementia [31]; no patient had sudden onset or stepwise decline of cognitive function suggestive of multi-infarct dementia that would confound subcortical-cortical clock-based analyses.

Aim 3

The third aim examined whether the two clock criteria would show expected differences in individuals with AD from those with frontal-subcortical disorders (PDD, VaD). The same 73 AD, 25 VaD, 18 PDD and 50 controls included in phase 2 were examined for command and copy differences per MoCA clock scoring criteria and the Cosentino criteria [10].

Clock Drawing Scoring

Both MoCA and Cosentino criteria require patients to 'draw the face of a clock with all the numbers present and set the hands to 10 after 11'.

Clock Scoring Criteria from the MoCA

Clock drawing accounts for 3 of the 30 total MoCA points [14] (10% of score). Scores range from 0 to 3 (3 = best) with points representing: (a) contour – the clock face is complete with only minor distortions (e.g. the circle is only slightly elongated or there is a small imperfection on closing the circle); (b) numbers – all numbers are present in the correct clockwise sequence; they must be

Table 1. Interrater ICC and 95% confidence intervals (in parentheses)

Set (n = 40)	Rater 1/rater 2	Rater 1/rater 3	Rater 2/rater 3	Across all 3 raters
MoCA clock criteria				
A	0.29 (-0.01 to 0.55)	0.12 (-0.19 to 0.42)	0.31 (0.01–0.57)	0.28 (0.10–0.48)
B	0.73 (0.55–0.85)	0.33 (0.03–0.58)	0.39 (0.09–0.62)	0.50 (0.22–0.71)
C	0.93 (0.88–0.96)	0.25 (-0.06 to 0.52)	0.23 (0.08–0.50)	0.46 (0.11–0.71)
D	0.82 (0.68–0.90)	0.52 (0.26–0.71)	0.29 (0.02–0.55)	0.58 (0.26–0.78)
Cosentino clock criteria				
A	0.92 (0.82–0.96)	0.92 (0.86–0.96)	0.92 (0.85–0.96)	0.92 (0.87–0.95)
B	0.91 (0.84–0.95)	0.82 (0.69–0.90)	0.83 (0.70–0.91)	0.86 (0.77–0.92)
C	0.92 (0.86–0.96)	0.86 (0.75–0.93)	0.83 (0.71–0.91)	0.87 (0.80–0.93)
D	0.97 (0.93–0.98)	0.96 (0.93–0.98)	0.86 (0.76–0.93)	0.89 (0.80–0.72)

Table 2. Intrarater reliability ICC and 95% confidence intervals (in parentheses)

	Rater 1	Rater 2	Rater 3
MoCA criteria			
Set A	0.24 (-0.08 to 0.51)	0.79 (0.63–0.88)	0.58 (0.33–0.75)
Set B	0.88 (0.79–0.94)	0.82 (0.68–0.90)	0.75 (0.58–0.86)
Cosentino criteria			
	0.99 (0.98–1.00)	0.95 (0.92–0.98)	0.95 (0.90–0.97)

located within their respective quadrants, and no numbers can be repeated; Roman numerals are acceptable, and numbers can be located outside of the clock face; (c) hands – both must be set to the correct time with the hour hand distinctively smaller than the minute hand, and both joining together near the center of the clock face.

Cosentino Clock Scoring Criteria

The Cosentino criteria [10] were empirically determined based on the classification of errors reported in previous studies of clock drawing in dementia [3, 5, 9, 11]. A more detailed description of the errors assessed (graphomotor dysregulation, perseveration, spatial, and time representation) is available for review [10].

Statistical Analysis

Inter- and intrarater reliability for all scores was assessed with one-way random, single-measure, intraclass correlation coefficients (ICC). Fisher's *r*-to-*Z* transformation examined differences in ICC *r* values. Confidence intervals (95%) are reported in tables. Discriminant function analyses examined classification rates. Mixed model ANOVAs examined hypothesized interactions between command and copy performances and group type (e.g. dementia, controls). Effect sizes for command to copy comparisons are reported with Cohen's *d* (small = 0.20, moderate = 0.50, large = 0.80 [32]). Clock scores were assessed for normality requirements and transformed if necessary. Alpha values were set at 0.01.

Results

Aim 1: Rater Reliability with MoCA and Cosentino Criteria

In table 1, *interrater reliability* for MoCA was low to moderate (interrater ICC range = 0.28–0.58). For Cosentino scoring criteria, high interrater reliability was achieved across and between rater pairs (ICC range = 0.86–0.92). MoCA interrater reliabilities were significantly lower than those with the Cosentino criteria (set A: $Z = -9.95$, $p < 0.001$; set B: $Z = -5.69$, $p < 0.001$; set C: $Z = -6.39$, $p < 0.001$; set D: $Z = -5.81$, $p < 0.001$).

In table 2, *intrarater reliability* for MoCA required 2 sets (ICC range for set A: 0.24–0.82; ICC range for set B: 0.83–0.92). Only 1 clock set was required to achieve high intrarater reliability for Cosentino criteria (ICC range set A: 0.95–0.99). Rater reliabilities for MoCA set B to Cosentino set A were statistically similar.

Examination of Command and Copy Conditions

Clocks of dementia patients (AD, VaD, PDD) were compared to those of controls who had little to no medical comorbidity (Charlson Comorbidity Index = 1.14 ± 1.44), had intellectual quotients in the average range (Wechsler Abbreviated Intelligence Scale IQ = 111.45 ± 12.98) and

Table 3. Mean \pm SD for demographic and basic cognitive measures of healthy controls (NC), all dementia participants, and dementia subtypes

	NC (n = 50)	Dementia (n = 231)	AD (n = 73)	IVD (n = 35)	PDD (n = 18)
Age, years	70.54 \pm 5.79	78.43 \pm 6.17	77.79 \pm 5.31	77.64 \pm 5.39	74.89 \pm 7.73
Education, years	15.55 \pm 3.23	11.95 \pm 2.93	12.44 \pm 2.78	11.60 \pm 2.70	11.67 \pm 4.07
M:F ratio	28:22	79:149	28:44	7:18	5:11 ¹
MMSE score	29.28 \pm 1.01	21.94 \pm 3.45	22.51 \pm 3.34	20.72 \pm 3.99	21.61 \pm 3.57
Junque	–	–	4.15 \pm 2.61	21.92 \pm 3.34	–

AD and ischemic vascular dementia (IVD) subtypes were screened for ‘mixed cortical and cortical dementia’ using clinical case conference and imaging analysis of vascular disease severity. Junque = Junque visual rating scale (reference) for white-matter abnormality severity (0–40 scale; 40 = maximum white-matter abnormalities). ¹ Missing data on 2 individuals with PPD with regard to sex.

had standardized neuropsychological scores assessing memory, language, visuospatial function, and in the average range (mean Z score range = –0.20 to 0.50). Relative to the controls, patients with dementia were older ($p < 0.001$), less educated ($p < 0.001$), and scored significantly lower on the MMSE ($p < 0.001$). Dementia subgroups were, however, similar in age [$F(2, 115) = 1.90, p = 0.155$], education [$F(2, 115) = 1.00, p = 0.373$] and MMSE [$F(2, 115) = 2.52, p = 0.085$]. Evidence of subcortical white-matter disease was low in AD (Junque score: AD = 4.15 \pm 2.61) and severe in VaD (Junque score: VaD = 21.92 \pm 3.34; table 3). Despite control and dementia group differences in age and education, there was no clinically significant correlation between clock performance and age (MoCA total error scores, all $r < -0.19$; Cosentino total error scores, all $r < 0.20$) or education (MoCA total error scores, all $r < 0.19$; Cosentino total error scores, all $r < -0.10$).

Aim 2: Dementia Classification with Command Condition Alone

The MoCA command condition correctly classified 43.7% of dementia (101/231) and 88.0% of control (44/50) participants, with 51.6% of the entire sample correctly classified. For Cosentino criteria, the command-only condition correctly predicted 57.1% of the dementia group (132/231) and 82.0% of control (41/50) participants, with 61.6% of the entire sample correctly classified.

For dementia subgroups, the MoCA command condition did not predict any AD or VaD patients; all were misclassified as controls or PDD (correct classification: AD = 0%, VaD = 0%; PDD = 61.1%, 11/18; normal controls = 88%, 44/50), with 33.1% of the entire sample correctly classified. For Cosentino criteria, each dementia sub-

group had some members correctly classified (correct classification: AD = 19.2%, 14/73; VaD = 40.0%, 10/25, PDD = 38.9%, 7/18; normal controls = 68.0%; 34/50], with 39.2% of the entire sample correctly classified.

Aim 3: Command and Copy Patterns for Controls and Dementia Subtypes

MoCA Criteria

There was a significant group (AD, VaD, PDD, normal controls) by condition interaction [$F(3, 162) = 5.88, p = 0.001$]. Planned pairwise comparisons showed significant command to copy improvement for AD ($p < 0.001, d = -1.26$) and VaD ($p = 0.013, d = -0.58$) but not PDD ($p = 0.205, d = -0.33$) or controls ($p = 0.032, d = -0.37$; fig. 1, 2; table 4).

Cosentino Criteria

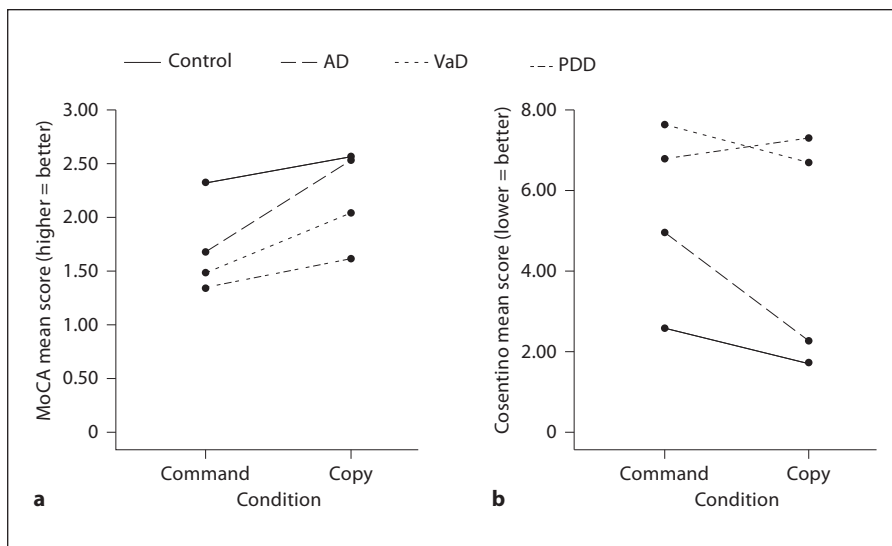
There was a significant group (AD, ischemic VaD, PDD, normal controls) by condition interaction [$F(3, 162) = 5.38, p = 0.001$]. Planned pairwise comparisons showed significant command to copy improvement for AD ($p < 0.001, d = 0.90$) but not VaD ($p = 0.16, d = 0.19$), PDD ($p = 0.70, d = -0.09$) or controls ($p = 0.06, d = 0.33$; fig. 1, 2).

Discussion

Clock drawing is an excellent screening tool for dementia, but test and scoring procedures vary. The current study confirmed that clock scoring methods impact rater reliability. Restricting clock drawing to a unitary command condition was not sufficient for diagnostic specificity for dementia or dementia subtype (AD, VaD or

Fig. 1. Control and dementia subgroup command and copy performances as scored with the MoCA criteria (a, from 0 to 3; 3 = perfect score) and Cosentino criteria (b, error ratings based on error subtypes; more errors = worse performance [10]).

Fig. 2. Example clocks from normal controls, AD, VaD and PDD patients. Scoring guide: normal controls, MoCA command = 3 (contour, numbers, hands); MoCA copy = 3 (contour, numbers, hands); Cosentino copy = 1 error (mental planning); AD, MoCA command = 1 (contour); MoCA copy = 3 (contour, numbers, hands); Cosentino command = 8 errors (gross motor, time representation, mental planning, perseveration and pull to stimulus); Cosentino copy = 3 errors (gross motor, mental planning); VaD, MoCA command = 2 (contour and numbers); MoCA copy = 2 (contour and numbers); Cosentino command = 7 errors (gross motor, time representation, mental planning, perseveration and pull to stimulus); Cosentino copy = 4 errors (time representation, mental planning); PDD, MoCA command = 1 (contour); MoCA copy = 1 (contour); Cosentino command = 12 errors (gross motor); Cosentino copy = 21 errors (gross motor, time representation, mental planning, perseveration and pull to stimulus).











Condition	Normal control	AD	VaD	PDD
Command	 MoCA: 3, Cosentino: 1	 MoCA: 1, Cosentino: 8	 MoCA: 2, Cosentino: 7	 MoCA: 1, Cosentino: 12
Copy	 MoCA: 3, Cosentino: 1	 MoCA: 3, Cosentino: 3	 MoCA: 2, Cosentino: 4	 MoCA: 1, Cosentino: 21

Table 4. Mean \pm SD for MoCA and Cosentino copy and command scores for healthy controls (NC), dementia patients and dementia subtypes

	NC (n = 50)	Dementia (n = 231)	Cortical/subcortical dementia subtypes		
			AD (n = 73)	VaD (n = 25)	PDD (n = 18)
MoCA scores					
Command	2.32 \pm 0.74	1.64 \pm 0.80	1.67 \pm 0.75	1.4 \pm 0.71	1.33 \pm 0.77
Copy	2.56 \pm 0.58	2.28 \pm 0.84	2.52 \pm 0.60	2.04 \pm 1.20	1.61 \pm 0.98
Cosentino scores					
Command	2.59 \pm 3.23	5.87 \pm 4.09	4.92 \pm 3.80	7.60 \pm 4.86	6.78 \pm 4.03
Copy	1.71 \pm 1.54	3.90 \pm 3.83	2.22 \pm 1.94	6.68 \pm 4.96	7.28 \pm 6.61

MoCA scores: higher scores = more intact; Cosentino scores: higher scores = more errors.

PDD). Inclusion of a copy condition demonstrated dementia subgroup differences.

Reliability was difficult to achieve with MoCA criteria. Although each of the 3 raters ultimately achieved reliability to self after training with 2 sets of clocks, only 2 of 3 raters achieved MoCA *interrater* reliability. The novice raters described MoCA criteria as 'vague', thereby leading to subjective interpretation. Retrospective MoCA score comparisons showed a rater variability of 1–3 points. Because the MoCA is a 30-point scale, examiner error can theoretically impact 10% of the total score. Scoring discrepancy could therefore impact whether a patient is scored in the impaired range (MoCA cutoff for impairment ≤ 25). Discrepancies of 2 points between clinicians/clinical assessments could impact treatment decisions. Rater errors resulting in a score *drop* from one clinical time point to the next could result in less than optimal clinical decision making. Scoring errors that *add* points may reduce the MoCA sensitivity and result in false impression of disease stability and subsequently a failure to treat. These clock scoring discrepancies warrant consideration for medical management consequences.

Raters immediately achieved high reliability using the Cosentino criteria. Scoring methods required preciseness (e.g. use of template to judge the circle face). The Cosentino criteria, however, required more time, i.e. raters reported taking up to 5 min to score 1 clock protocol (command plus copy), whereas the MoCA clock scoring typically took less than 1 min to score. Overall, if one uses the MoCA clock scoring for clinical use or research comparison across sites, it appears prudent to require training and practice. This will establish reliability before implementing the test clinically. Otherwise, we encourage more objective error-based methods such as the Cosentino method.

Regardless of which clock scoring method was used (MoCA vs. Cosentino), drawing a clock to the command condition alone could not effectively distinguish demented individuals from controls; groups were distinguished *at chance levels*. Moreover, simple clock drawing to command could *not effectively distinguish dementia subgroups*. Although the command condition with Cosentino criteria was slightly improved over MoCA, accuracy was still quite low.

Diagnostic differences occurred only when command and copy performances were compared. AD patients improved from command to copy such that by the copy condition, this group performed similarly to the control group. This was observed for both MoCA and Cosentino

clock scoring criteria. For the subcortical dementia patients, PDD patients failed to improve on copy with either scoring technique. VaD patients failed to improve with copy when their clocks were scored according to Cosentino criteria, but showed marginal improvement with the MoCA criteria. This discrepancy may reflect the increased sensitivity of classic neuropsychological error score analysis. Error analysis for perseverations, motor errors, time representation errors and spatial planning errors across the command and copy conditions promotes consideration of underlying cognitive constructs (i.e. semantic memory, executive control) which are differentially impaired in dementia (for review, see historically relevant works from neurology and neuropsychology [3, 6]). The control group performed well on both command and copy (largely at ceiling; no errors), and although there was no statistical difference, there was a trend for command to copy improvement associated with a moderate effect size. Little to no errors with subtle improvement in command to copy can be considered an expectation for normal performance. This corresponds to previous reports [12].

There is clear assessment value for incorporating both command and copy conditions [5, 9, 11]. Drawing to command requires intact auditory comprehension, the ability to persist in drawing, remembering instructions (e.g. working memory for drawing the clock face, hands and numbers), verbal and visual memory, as well as spatial planning (numbers, hand placement, page placement) [3, 5, 9]. When executing a drawing, the examiner's command must be understood and particularly the syntactic comprehension of '10 after 11' (see other versions, as well [4]). Hence, a poor command score can result for many reasons. This explains the general insensitivity of the overall command score and its poor discriminability for dementia classification in our sample. However, drawing a clock to a model is specifically related to visual and executive function [5, 10]. There is less demand on memory, but an increased need for visuospatial integration as well as inhibitory functions. Individuals with deficits in these areas (e.g. PDD, VaD) often transfer errors from the command condition to that of the copy condition. Indeed, the production of copy errors associates with worse mental flexibility, inhibitory functions and planning abilities. There is, however, little to no relationship between copy errors and declarative and semantic memory functions [5, 10, 11] which are often compromised in AD. This command-copy dissociation is consequently useful for assessing patients' cognitive strengths and

weaknesses and assisting with preliminary differential diagnosis.

Regarding limitations, this study was based on acquired data from clinical and research populations. All sources, however, used the same instructions and were administered by clinical neuropsychologists. Study strengths include consideration for a clock approach commonly used on inpatient floors by new residents (e.g. MoCA) and clinicians not familiar with dementia subtypes. Our novice raters resembled these types of providers. Findings therefore expand upon recent studies reporting clock scoring reliability among experienced clinicians [33].

Overall, the clock drawing test is an efficient technique for assessing numerous cognitive functions. Based on our collective findings, we provide three recommendations. First, regardless of which clock scoring method is chosen (i.e. rapid MoCA versus more error-based paradigms), clinicians should practice for reliability. Second, we caution making clinical judgments based on MoCA change scores that are acquired from another provider or clinician with whom reliability has not been established. Third, should the MoCA be used clinically, we strongly recommend supplementing the test with a clock copy af-

ter asking the patient to complete the command condition. This often takes less than 1 min to administer and will provide additional clinical information.

In summary, this investigation identified limitations to clock drawing approaches in one modern dementia screening test, the MoCA. We provide recommendations to enhance the clinical value of the test. We encourage further investigations on rater training for the MoCA clock test item and variations in overall MoCA scores to improve this test's value within community and inpatient settings.

Acknowledgment

We are grateful to the National Institute for Neurological Disorders and Stroke for providing the Recovery Act Student Summer Supplement that initiated this project (K23NS060660-2S1; to C.P.) and supported it in part.

Disclosure Statement

There are no conflicts of interest to disclose.

References

- 1 Kleist K: Der Gang und der gegenwärtige Stand der Apraxieforschung. *Neurol Psychiatr* 1912;1:342–352.
- 2 Head H: *Aphasia and Kindred Disorders of Speech*. New York, Macmillan, 1926.
- 3 Kaplan E: A process approach to neuropsychological assessment; in Boll T, Bryant BK (eds): *Clinical Neuropsychology and Brain Function: Research, Measurement, and Practice*. Washington, American Psychological Association, 1988.
- 4 Shulman KI: Clock-drawing: is it the ideal cognitive screening test? *Int J Geriatr Psychiatry* 2000;15:548–561.
- 5 Libon DJ, Malamut BL, Swenson MR, Cloud BS: Further analysis of clock drawings among demented and non-demented subjects. *Arch Clin Neuropsychol* 1996;11:193–211.
- 6 Goodglass H, Kaplan E: *The Assessment of Aphasia and Related Disorders*. Philadelphia, Lea & Febiger, 1983.
- 7 Ismail, Z, Rajji TK, Shulman KI: Brief cognitive screening instruments: an update. *Int J Geriatr Psychiatry* 2010;25:111–120.
- 8 Libon DJ, Bogdanoff B, Bonavita J, Skalina S, Cloud BS, Resh R, Cass P, Ball SK: Dementia associated with periventricular and deep white matter alterations: a subtype of subcortical dementia. *Arch Clin Neuropsychol* 1997;12:239–250.
- 9 Libon DJ, Swenson R, Baronski EJ, Sands LP: Clock drawing as an assessment tool for dementia. *Arch Clin Neuropsychol* 1993;8:405–415.
- 10 Cosentino S, Jefferson A, Chute DL, Kaplan E, Libon DJ: Clock drawing errors in dementia: neuropsychological and neuroanatomical considerations. *Cogn Behav Neurol* 2004;17:74–84.
- 11 Rouleau I, Salmon DP, Butters N, Kennedy C, McGuire K: Quantitative and qualitative analyses of clock drawings in Alzheimer's and Huntington's disease. *Brain Cogn* 1992;18:70–87.
- 12 Cacho J, Garcia-Garcia R, Fernandez-Calvo B, Gamazo S, Rodriguez-Perez R, Almeida A, Contador I: Improvement pattern in the clock drawing test in early Alzheimer's disease. *Eur Neurol* 2005;53:140–145.
- 13 Nagahama Y, Okina T, Suzuki N, Nabatame H, Matsuda M: Neural correlates of impaired performance on the clock drawing test in Alzheimer's disease. *Dement Geriatr Cogn Disord* 2005;19:390–396.
- 14 Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H: The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005;53:695–699.
- 15 Folstein MF, Folstein SE, McHugh PR: "Mini-mental state": a practical method for grading the clinician. *J Psychiatr Res* 1975;12:129–138.
- 16 Dalrymple-Alford JC, MacAskill MR, Nakas CT, Livingston L, Graham C, Crucian GP, Melzer TR, Kirwan J, Keenan R, Wells S, Porter RJ, Watts R, Anderson TJ: The MoCA: well-suited screen for cognitive impairment in Parkinson disease. *Neurology* 2010;75:1717–1725.
- 17 Blair M, Kertesz A, McMonagle P, Davidson W, Bodi N: Quantitative and qualitative analyses of clock drawing in frontotemporal dementia and Alzheimer's disease. *J Int Neuropsychol Soc* 2006;12:159–165.

- 18 Gelb DJ, Oliver E, Gilman S: Diagnostic criteria for Parkinson disease. *Arch Neurol* 1999;56:33–39.
- 19 Emre M, Aarsland D, Brown R, Burn DJ, Duyckaerts C, Mizuno Y, Broe GA, Cummings J, Dickson DW, Gauthier S, Goldman J, Goetz C, Korczyn A, Lees A, Levy R, Litvan I, McKeith I, Olanow W, Poewe W, Quinn N, Sampaio C, Tolosa E, Dubois B: Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Mov Disord* 2007;22:1689–1707, quiz 1837.
- 20 McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM: Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology* 1984;34:939–944.
- 21 Chui HC, Victoroff JI, Margolin D, Jagust W, Shankle R, Katzman R: Criteria for the diagnosis of ischemic vascular dementia proposed by the State of California Alzheimer's Disease Diagnostic and Treatment Centers. *Neurology* 1992;42:473–480.
- 22 APA: Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), ed 4. Washington, American Psychiatric Association, 1994.
- 23 Charlson ME, Pompei P, Ales KL, MacKenzie CR: A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–383.
- 24 Wechsler D: Wechsler Abbreviated Scale of Intelligence. San Antonio, Psychological Corporation, 1999.
- 25 Wiederkehr S, Simard M, Fortin C, van Reekum R: Validity of the clinical diagnostic criteria for vascular dementia: a critical review. Part II. *J Neuropsychiatry Clin Neurosci* 2008;20:162–177.
- 26 O'Brien JT, Erkinjuntti T, Reisberg B, Roman G, Sawada T, Pantoni L, Bowler JV, Ballard C, De Carli C, Gorelick PB, Rockwood K, Burns A, Gauthier S, De Kosky ST: Vascular cognitive impairment. *Lancet Neurol* 2003;2:89–98.
- 27 Junque C, Pujol J, Vendrell P, Bruna O, Jodar M, Ribas JC, Vinas J, Capdevila A, Marti-Vilalta JL: Leuko-araiosis on magnetic resonance imaging and speed of mental processing. *Arch Neurol* 1990;47:151–156.
- 28 Price CC, Garrett KD, Jefferson AL, Cosentino S, Tanner JJ, Penney DL, Swenson R, Giovannetti T, Bettcher BM, Libon DJ: Leukoaraiosis severity and list-learning in dementia. *Clin Neuropsychol* 2009;23:944–961.
- 29 Price CC, Jefferson AL, Merino JG, Heilman KM, Libon DJ: Subcortical vascular dementia: integrating neuropsychological and neuroradiologic data. *Neurology* 2005;65:376–382.
- 30 Libon DJ, Price CC, Giovannetti T, Swenson R, Bettcher BM, Heilman KM, Pennisi A: Linking MRI hyperintensities with patterns of neuropsychological impairment: evidence for a threshold effect. *Stroke* 2008;39:806–813.
- 31 Rosen TJ: Cortical vs subcortical dementia: neuropsychological similarities. *Arch Neurol* 1987;44:131.
- 32 Cohen J: *Statistical Power Analysis for the Behavioral Sciences*, ed 2. Hillsdale, Erlbaum Associates, 1988.
- 33 Nair AK, Gavett BE, Damman M, Dekker W, Green RC, Mandel A, Auerbach S, Steinberg E, Hubbard EJ, Jefferson A, Stern RA: Clock drawing test ratings by dementia specialists: interrater reliability and diagnostic accuracy. *J Neuropsychiatry Clin Neurosci* 2010;22:85–92.