CLINICAL RESEARCH

Reliability and Validity of the AB Clock Test (ABCT): An Objective Approach to Scoring the Clock Drawing Test

Jamil Kanji, BSc (Hons.)
D.William Molloy, MB, MRCPI, FRCPC
Timothy Standish, MA
David Lewis, PhD
Roy Chivers, MD, PhD, CCFP

ABSTRACT

The AB Clock Test (ABCT) has objective scoring criteria incorporating a template. The reliability and validity of the ABCT was compared to the Sunderland and Watson scoring methods. Twenty-six patients, referred to an outpatient clinic for memory problems, were administered the clock drawing test as well as the Standardized Mini-Mental State Examination (SMMSE) and the AB Cognitive Screen (ABCS). Five raters independently scored the 26 clocks using each of the three scoring methods. Reliability of the Sunderland method was 0.95 followed by the ABCT (0.94) and the Watson (0.85) techniques. Convergent validity was measured by bivariate correlations between the three methods, the SMMSE and the ABCS. Correlations were highest with the ABCT, followed by Sunderland and the Watson. The Sunderland method correlated highly with a diagnosis of normal cognition, mild cognitive impairment or dementia, followed by the ABCT and the Watson. The ABCT demonstrates reliability comparable to the Sunderland method while validity was better than the Sunderland and Watson methods.

INTRODUCTION

A wide variety of cognitive screening tests are currently used to assess patients who present with cognitive impairment and memory loss. One of these tests, the clock drawing test (CDT), is widely used to screen people who present with memory loss/cognitive impairment. While the test is easy to administer and requires little training, there are a wide variety of different CDT administration practices and scoring guidelines currently used to administer and score this test.[1]

The CDT meets the defined criteria for a successful cognitive screening instrument.[2] These criteria include: (i) taking less than 2 minutes to administer; (ii) accepted by patients; (iii) simple to score; (iv) independent of educational and cultural backgrounds; (v) good inter-rater and test re-test reliabilities; (vi) a high level of sensitivity as well as specificity; (vii) coexisting validity; and (viii) predictive validity. The CDT tests a wide range of cognitive abilities, is fast to administer, and is well tolerated by patients. It tests comprehension, planning, visual memory, and reconstruction.[3] This simple test also evaluates patients' visuospatial abilities, motor programming, motor execution, numerical knowledge, semantic instruction, and the inclination to interpret a command literally.[3]

We have developed standardized administration and scoring guidelines for a variety of different cognitive tests.[4,5,6] Because of the wide variability administration and scoring guidelines used for clock drawing tests, we developed clear and explicit guidelines for administration and scoring, which we term the AB Clock Test (ABCT). The ABCT uses a plastic transparent sheet with a vertical and horizontal axis and four quadrants to objectively measure the positioning of the hands and numbers on the clock. The method is part of a package known as the *abcdes Short Comprehensive Screen*, which is used to quantify cognitive deficits in patients who present with memory problems and potential cognitive impairment.[4]

In this study, we examined and compared the convergent validity and inter-rater reliability of the ABCT with the scoring methods developed by Sunderland, *et al.*[7] and Watson, *et al.*[8], hereafter referred to as the Sunderland and Watson techniques, respectively. Among the many clock scoring methods available, the Sunderland scale was chosen for comparison because it is widely used and accepted, and has demonstrated high reliability. Like the ABCT, it is also administered without the provision of a pre-drawn circle (see Table 1). The Watson method of scoring was chosen as the other comparator because it represents a very objective approach to scoring the CDT (described below) and also uses a pre-drawn circle. The administration of the ABCT is no different from these, however we believe its scoring system, using the template, may make it more sensitive to slight changes in test-retest situations that are an important tool in diagnosing cognitive disorders. This hypothesis will be

investigated in another paper. First we need to establish that the convergent validity and inter-rater reliability of the ABCT are at least equivalent or better than those of Sunderland or Watson.

METHODS

The data used in this study was collected retrospectively from charts of patients seen from January 6 – April 14, 2003. The subjects in this study were patients who attended the geriatric outpatient Memory Clinic at St. Peter's Hospital in Hamilton, Ontario, a university-affiliated teaching hospital in a large urban center. Patients had originally been referred to our service by physicians to assess problems with memory and other signs of functional decline. A trained geriatric team collected initial patient data using a standardized geriatric assessment.[9] Trained geriatricians gave a diagnosis of Alzheimer's dementia (AD) using the NINCDS-ADRDA criteria,[10] mild cognitive impairment (MCI), or no cognitive impairment (normal). Cognitive impairment was measured using the Standardized Mini-Mental Status Examination (SMMSE),[4] Global Deterioration Score (GDS),[11] and the AB Cognitive Screen (ABCS).[12]

The CDT was administered to each patient using the rules established for the ABCT (Table 1) as part of the above described assessment. The clocks from twenty-six patients were chosen by a non-rater individual to portray a variety of clock presentations to highlight scoring anomalies of the different scoring techniques. Each clock was scored by five different raters using the ABCT, Sunderland, and Watson methods. The raters were provided with standard printed instructions on how to score each scale (Tables 1-3, Figure 1).

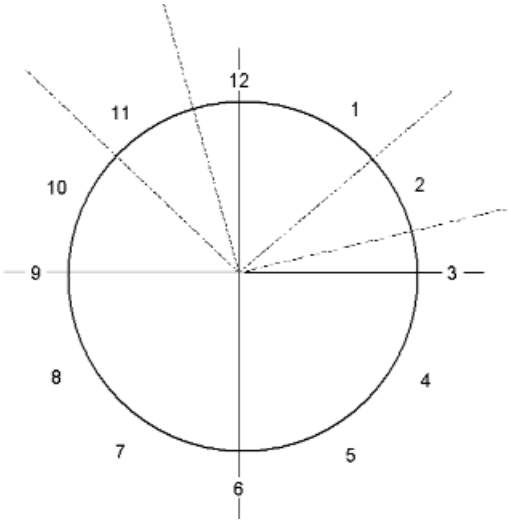

Figure 1. Transparency scoring template used in scoring of the CDT using the ABC Clock Test method.

The Watson method of scoring the CDT divides the clock into four quadrants. This technique does not score the hands, but focuses exclusively on the placement of numbers or symbols within each quadrant. The instructions provided to the raters for the Watson technique are outlined in Table 2. To avoid marking on the clocks, we provided raters with a transparency with pre-drawn perpendicular lines for the Watson technique.

The Sunderland scoring technique provides the rater with examples of clocks and written descriptions with which to compare. Each example clock has an associated score. The full instructions provided to the raters for this method are shown in Table 3.

To score the clocks using the ABCT technique, raters were provided with the transparency to overlay on the clocks (Figure 1) and the instructions shown in Table 1.

Raters included staff from hospital clinics, medical students, and administrative services who had no prior cognitive testing experience. Raters were blinded to one another and to the identities, demographics, diagnoses, SMMSE, GDS, and ABCS scores of the patients. Each rater scored all 26 clocks using one method before switching to another method. To minimize bias or a tendency towards harmonization between an individual rater's scores among the three methods for a particular clock, raters were not given access to their scores from the other methods.

Statistical analysis was conducted using SPSS (version 11.5). To maintain consistency between methods, the Watson scores were recoded so that a higher score indicated greater impairment; each of the raters' scores was then recoded on a scale from 0 to 100. We assessed the alpha reliability using standardized intraclass correlations. We also examined partial correlations to identify areas of inconsistent scoring.

Convergent validity was assessed by computing the mean scores by the five raters for each method and running partial correlations against both the ABCS and the SMMSE, on the premise that a good clock drawing method should correlate strongly, but not perfectly. Both the ABCS and SMMSE have already been validated as tests of cognitive function.[5,12] Similarly, we assessed their concurrent validity by analyzing the capacity of each of the three averages to distinguish between diagnoses of AD, MCI, and normal cognition, using one-way analysis of variance (ANOVA).

RESULTS

A total of 26 patients contributed data to the analyses. A description of the sample of patients can be found in Table 4. In general, traits varied as expected by diagnosis, except that those with normal cognition were slightly older than the other two groups (not significant), and scored slightly lower on the SMMSE than those with MCI (27.7 vs 28.3) ($F = 12.1$, $p < 0.001$). The SMMSE does not reliably discriminate between normal cognition and mild cognitive impairment.[17, 18]

Table 5 shows the intraclass correlations for the three methods. Correlation for the Sunderland was 0.95, the ABCT 0.94 and the Watson 0.85. Table 6 shows bivariate correlations between the raters (with rater 1 treated as the 'dependent') for each method in turn. Correlations ranged to a low of 0.313 for rater number 4 using the Watson method. Table 7 demonstrates the bivariate correlations between the three methods with the SMMSE and ABCS (clock domain excluded), as an indicator of convergent validity. The correlations with the SMMSE

Table 1. AB Clock Test (ABCT) procedure[4]

Administration

1. Give a blank sheet of paper and pencil to the patient.
2. Gesture with your finger in a large circle while saying "Draw a large circle here please."
3. When the circle is drawn, say "Now put in the numbers like the face of a clock."
4. Then say, "Set the hands to show ten past eleven."
5. You may prompt at each stage... "put in the numbers...put the time as ten past eleven."
6. Note, if the person draws a small clock face, ask him/her to draw it again because he/she will not be able to put in the numbers and it will be difficult to score.

Scoring

Place the scoring template (Figure 1) over the completed clock with the template's "12 o'clock" line placed over the subject's 12. Adjust the template to maximize the score for the numbers and hands. Record scores as follows.

Numbers

- For the numbers 12, 3, 6, and 9 score two (2) points if they touch their respective lines, one (1) point if missed, and zero (0) if the number is omitted.
- For the numbers 1, 2, 4, 5, 7, 8, 10, and 11 score two (2) points for each number in the correct quadrant, one (1) point if the number is outside the quadrant, and zero (0) if the number is omitted.
- Subtract one point for each number repeated.

Hands

Score the placement of the entire hand. Each hand is scored separately.

- If the hands are drawn within range, score three (3) points.
- If the hands are touching the hatched line, score two (2) points.
- If the hands are drawn outside the hatched line, score one (1) point.
- If the hands are omitted, score zero (0).

Deduct one (1) point if the hands do not join at the pivot.

Maximum score is 30.

Table 2. Scoring the CDT according to the Watson method[8]

1. Draw one line through the center of the circle and the number 12 or mark that best corresponds to the number 12. Draw another line perpendicular to this line also through the middle of the circle. This will divide the clock into four segments.
2. Moving in a clockwise fashion, count the number of digits in each quadrant beginning with the number 12. If a digit falls on one of the quadrant lines, it is included in the quadrant preceding the line. A total of 3 digits in one quadrant is considered correct.
3. If there is an error in numbers of digits in the first three quadrants (spanning numbers 12 through 9), assign a score of 1. For any error in the number of digits in the last quadrant, assign a score of 4.
4. The sum of the scores for each quadrant is the total score for the clock.

Table 3. The Sunderland method for scoring the CDT[7]

SCORE	DESCRIPTION
10-6	**Drawing of clock face with circle and numbers is generally intact.**
10	Hands are in correct position.
9	Slight errors in placement of hands.
8	More noticeable errors in placement of hour and minute hands.
7	Placement of hands is significantly off course.
6	Inappropriate use of clock hands (ie. Use of digital display or circling of numbers despite repeated instructions.)
5-1	**Drawing of a clock face with circle and numbers is not intact.**
5	Crowding of numbers at one end of the clock or reversal of numbers. Hands may still be present in some fashion.
4	Further distortion of number sequence. Integrity of clock face is now gone (ie. numbers missing or placed at outside of the boundaries of the clock face).
3	Numbers and clock face no longer obviously connected in drawing. Hands are not present.
2	Drawing reveals some evidence of instructions being received but only a vague representation of a clock.
1	Either no attempt or an uninterpretable effort is made.

Table 4. Description of sample

	Overall	Type of Dementia AD	MCI	Normal
N	26	14	6	6
Age (years)				
Mean; SD	74.8; 9.0	74.4; 9.1	72.3; 6.9	78.1; 11.2
(range)	(56-98)	(56-93)	(61-81)	(66-98)
Percent Male	34.6	21.2	50.0	50.0
Education (years)				
Mean; SD	11.7; 2.7	11.6; 2.8	11.7; 3.4	11.7; 3.4
English as Second Language (%)	11.5	14.2	0	16.7
Memory Loss (%)	76.9	100	100	0
Duration of Memory Loss (years)				
Mean; SD	4.2; 2.9	3.4; 1.6	5.7; 4.3	N/A
(range)	(0-12)	(1-6)	(1-12)	
SMMSE Score				
Mean; SD	24.6; 4.5	21.7; 3.9	28.3; 1.5	27.7; 2.7
(range)	(13-30)	(13-26)	(26-30)	(23-30)
GDS Score (Reisberg)				
Mean; SD	3.4; 1.6	4.5; 0.8	3.2; 1.0	1.2; 0.4
(range)	(1-6)	(3-6)	(2-5)	(1-2)
ABCS (excluding clock)				
Mean; SD	66.5; 21.3	52.0; 18.4	82.2; 9.2	84.5; 5.8
ABCS (including clock)				
Mean; SD	92.9; 25.5	76.4; 23.9	110.5; 8.5	113.8; 5.4

Scores are expressed as means and standard deviations.

range from 0.89 (ABCT) to 0.56 (Watson). Correlations with the ABCS range from 0.88 (ABCT) to 0.59 (Watson).

Table 8 shows the results of a one-way ANOVA of the three averages with the diagnostic group dependent. It indicated the probability of error within the groups, which was low with the Sunderland and ABCT (p=0.001 and 0.006 respectively) and higher with the Watson (p=0.227).

DISCUSSION

The intent of the current study was to compare the ABCT technique of scoring the CDT with the Sunderland and Watson CDT scoring scales in terms of inter rater reliability and correlation with diagnosis of cognitive impairment. The choice as to which scoring scale a clinician should adopt is generally dependent upon the specific needs and skills of the clinician.[1]

In the present study, we found high correlation coefficients for all of the CDT scoring techniques examined. The highest agreement was found between the ABCT and Sunderland methods. The Watson performed the worst in terms of heterogeneous scores and ABCT the best. This indicates that the same patient assessed by the same rater is more likely to be found demented by the Watson method than on the other two and least likely by the ABCT.

Further examination of variability for scores obtained by different raters with the same scale was carried out. Analysis of partial correlations yielded correlation values that were quite close for both the ABCT and Sunderland technique. However, when one looks at the data for the Watson partial correlations, there is a wide variability between individual raters with correlations ranging from 0.31 up to 0.72.

These differences may be explained by the fact that both the ABCT and Sunderland methods administer the test without the provision of a pre-drawn circle and incorporate patients' clock hands as a component of the overall score. The Watson method uses a pre-drawn circle and excludes clock hands in the scoring procedure.

Similar to previous studies such as those of Juby[13] and Brodaty and Moore[14] which compared various CDT scoring techniques, we found strong correlations between the CDT and other cognitive tests like the SMMSE and ABCS. The ABCT had the highest correlations with the SMMSE and ABCS tests.

There were some weaknesses in this study. The sample size

of 26 clocks is small. A larger sample would have likely decreased variability. There are some slight differences in administration as well as scoring among the three methods e.g. the Watson specifies a pre-drawn circle, but all subjects were instructed to draw their clocks according to the ABCT instructions. We believe this difference in administration is small, and that the template for the ABCT scoring technique compensates for the circle drawn by the patient.

Table 5. Intraclass correlations of clock scoring methods

Scale	ABCT	Sunderland	Watson	All
ICC	0.94	0.95	0.85	0.94

This table shows the variability in scores for all clocks by all raters for each method.

Table 6. Partial correlations by rater

	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
ABCT	1.000	0.776	0.797	0.782	0.671
Sunderland	1.000	0.799	0.888	0.879	0.744
Watson	1.000	0.617	0.760	0.313	0.715

All correlations are $p < 0.001$
This table illustrates a subsection of the complete correlation matrix. It indicates the variations between different raters. Rater 1 is chosen as the dependent variable.

Table 7. Bivariate correlations of clock scoring means against SMMSE and ABCS (without clock domain)

	SMMSE	ABCS
ABCT	0.89	0.88
SUNDERLAND	0.83	0.86
WATSON	0.56	0.59

All correlations are $p < 0.001$

Table 8. One-way ANOVA, clock scores by diagnosis group

		Sum of Squares	*F*	*p*
ABCT	Between Groups	2170.6	6.6	0.006
	Within Groups	3811.1		
SUNDERLAND	Between Groups	8568.4	15.5	0.001
	Within Groups	6358.1		
WATSON	Between Groups	2565.5	1.6	0.227
	Within Groups	18666.9		

The ABCT shows a level of inter rater reliability comparable to that of the established Sunderland method, while having some advantages over the other techniques. First, it is very simple to teach. The five raters were given printed instructions in scoring and achieved high inter rater reliability without the need for formal training. Compared to the Sunderland technique, the ABCT method provides for more objectivity and reduces variability introduced by different notions of a correct clock or proper placement of numbers/hands by raters. The wording of the scoring scale developed for the Sunderland technique leaves itself open to subjective interpretation. For example, for a score of 9, the descriptor reads: "Slight error in placement of hands." The interpretation of "slight error" varied from scorer to scorer. The Sunderland method is presented to the rater with predescribed clocks for each score. In spite of these potential limitations, the Sunderland method is very reliable.

Overall the CDT is useful in screening populations for dementia and/or cognitive impairment. It requires only a paper and pen and is easy to administer with the various scales that have been developed. The CDT can also be administered successfully to the hearing impaired, very ill, and also those who may be poorly educated.[15] The ABCT demonstrated comparable inter rater reliability and validity to the Sunderland and was better than the Watson method. We were surprised that the Watson did not demonstrate higher reliability, given that its scoring method seems to be objective and its original published inter-rater reliability was 0.90 - 0.93.[8] A comparison study of five clock scoring methods also resulted in reliability of 0.90.[16]

Perhaps the most interesting finding in this study was that the ABCT had greater validity with cognitive scores. The ABCT had a higher correlation with SMMSE and ABCS scores than Sunderland, while Sunderland had greater validity with clinical diagnosis. Discussions among the research group and the raters after the end of the study revealed some ways in which the ABCT training and instruction might be improved, to increase the inter rater reliability even more. This study did not address the sensitivity of the three methods to changes in cognition with serial testing. It is our expectation that the scoring template will provide its greatest benefit in detecting small changes in clocks which may correlate with small cognitive changes. Additional study is also needed to evaluate the sensitivity of the ABCT with serial testing, in a larger sample size and also its overall effectiveness when combined with the ABCS to evaluate cognitive function overall.

ACKNOWLEDGEMENTS

This work was supported by funding from the Scottish Rite Charitable Foundation of Canada and Eli Lilly Canada Inc.

REFERENCES

1. Tuokko, H, Hadjistavropoulous, T, Rae, S. et al. (2000). A comparison of alternative approaches to the scoring of clock drawing. *Arch Clin Neuropsychol.* 15:137-48.
2. Shulman, KI. (2000). Clock-drawing: is it the ideal cognitive screening test? *Int J Geriatr Psychiatry.* 15:548-61.
3. Royall, DR, Cordes, JA, Polk, M. (1998). CLOX: an executive clock drawing task. *J Neurol Neurosurg Psychiatry*. 64:588-94.
4. Molloy, DW. (2003). *abcdes administration and scoring guidelines: a practical guide to the abcdes short comprehensive screen.* Dundas: Newgrange Press Ltd.
5. Molloy, DW, Alemayehu, E, Roberts, R. (1991). Reliability of a standardized mini-mental state examination compared with the traditional mini-mental state examination. *Am J Psychiatry.* 148:102-5.
6. Standish, TI, Molloy, DW, Bedard, M et al. (1996). Improved reliability of the standardized Alzheimer's Disease Assessment Scale (SADAS) compared with the Alzheimer's Disease Assessment Scale (ADAS). *J American Geriatr Soc.* 44:712-6.
7. Sunderland, T, Hill, JL, Mellow, AM et al. (1989). Clock drawing in Alzheimer's disease: a novel measure of dementia severity. *J American Geriatr Soc*. 37:725-9.
8. Watson, YI, Arfken, CL, Birge, SJ. (1993). Clock completion: an objective screening test for dementia. *J American Geriatr Soc*. 41:1235-40.
9. Lever, J, Molloy, DW, Vertesi, A et al. (2000). Developing an integrated geriatric assessment database and protocol. *Mature Medicine Canada*. Oct/November:3-4.
10. McKhann, G, Drachman, D, Folstein, M et al. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology.* 34:939-44.
11. Reisberg, B, Ferris, SH, de Leon, MJ et al. (1982). The global deterioration scale for assessment of primary degenerative dementia. *Am J Psychiatry*. 139:136-9.
12. Molloy, DW, Standish, TI, Lewis, DL. (2005). Screening for mild cognitive impairment: comparing the SMMSE and the ABCS. *Can J Psychiatry*. 50:52-8.
13. Juby, A. (1999). Correlation between the Folstein Mini Mental State Examination and three methods of clock drawing scoring. *J Geriatr Psychiatry Neurol*. 12:87-91.
14. Brodaty, H, Moore, CM. (1997). The clock drawing test for dementia of the Alzheimer's type: a comparison of three scoring methods in a memory disorders clinic. *Int J Geriatr Psychiatry.* 12:619-27.
15. Death, J, Douglas, A, Kenny, RA. (1993). Comparison of clock drawing with Mini-Mental State Examination as a screening test in elderly acute hospital admissions. *Postgrad Med J*. 69:696-700.
16. Schramm, U, Berger, G, Muller, R et al. (2002). Psychometric properties of Clock Drawing Test and MMSE or Short Perfomance Test (SKT) in dementia screening in a memory clinic population. *Int J Geriatr Psychiatry.* 17**:**254-60.
17. Tang-Wai, D, Knopman D, Geda Y et al. (2003). Comparison of the short test of mental status and the Mini-Mental State Examination in mild cognitive impairment. *Arch Neurol*. 60:1777-81.
18. Karrasch, M, Sinerva, E, Gronholm, P et al. (2005). CERAD test performances in amnestic mild cognitive impairment and Alzheimer's disease. *Acta Neurol Scand*. 111:172-79.

Author Biographies

Jamil Kanji is a third-year medical student at McMaster University who studied biochemistry at Queen's University and plans to pursue a career in internal medicine.

Dr. W. Molloy is a professor of medicine at McMaster University, consultant geriatrician at St. Peter's Hospital and Hamilton Health Sciences, and holder of the St. Peter's McMaster Chair in Aging. He completed medicine at University College in Cork, Ireland and trained in Geriatrics at the University of Manitoba, University of Western Ontario, and McMaster University.

Timothy Standish is a Research Coordinator at the St. Peter's Centre for Studies in Aging and completed his Master's degree in Political Science at McMaster University.

Dr. David Lewis is a faculty member and research scientist for the Faculty of Health Sciences, McMaster University, St. Joseph's Healthcare, and evaluator and researcher with the Regional Geriatrics Program of Hamilton, Ontario. He has completed his doctorate in Sociology.

Dr. Roy Chivers is the medical director of two nursing homes in Haldimand Norfolk and Inquest Coroner for the Province of Ontario. He completed medicine at Queen's University and holds a doctorate in Gross Human Anatomy and General Pathology from the University of Manitoba.