Diagnostic Performance of Clock Drawing Test by CLOX in an Asian Chinese Population

Philip Lin-Kiat Yap\textsuperscript{a, c} Tze-Pin Ng\textsuperscript{a, b} Mathew Niti\textsuperscript{a, b} Donald Yeo\textsuperscript{a, e} Linda Henderson\textsuperscript{d}

\textsuperscript{a}Gerontological Research Programme, Faculty of Medicine, and \textsuperscript{b}Department of Psychological Medicine, National University of Singapore, \textsuperscript{c}Department of Geriatric Medicine and \textsuperscript{d}Department of Nursing, Alexandra Hospital, and \textsuperscript{e}Brain Centre, Singapore General Hospital, Singapore, Singapore

Key Words
Clock Drawing Test · CLOX · Cognitive screen · Executive function, dementia

Abstract
Background/Aims: Clock Drawing Tests are commonly used for cognitive screening, but their clinical utility has not yet been studied in Chinese Singaporeans. We examined the usefulness of a Clock Drawing Test, CLOX, in detecting dementia in our population and explored its performance in the dementia subtypes, Alzheimer’s disease (AD), and the vascular composite group (VCG) of AD with cerebrovascular disease and vascular dementia. Method: CLOX was administered to 73 subjects (49.3%) with dementia and 75 healthy controls (50.7%). Receiver operating characteristic analysis determined the diagnostic accuracy and optimal cut-off scores, stratified by education. Analysis of Variance was used to compare CLOX scores between AD and VCG. Results: The diagnostic accuracy (area under the curve) was 84 and 85% for CLOX1 and CLOX2, respectively. Cut-offs at 10 for CLOX1 and 12 for CLOX2 yielded sensitivities of 75.3 and 75%, and specificities of 76 and 80%, respectively. The mean CLOX1 but not CLOX2 scores for AD (8.1) and VCG (5.5) remained significantly different (p = 0.002) after adjustment for the covariates age, gender, education, MMSE and dementia stage. Conclusion: Our results support CLOX as a valid cognitive screen in Singaporean Chinese with adequate psychometric properties. In addition, CLOX may aid as an adjunct in differentiating AD from dementia with a vascular element, e.g. AD with cerebrovascular disease and vascular dementia.
translated versions. The problem of language and education is even more pronounced in a multi-ethnic and multi-dialect Asian country like Singapore, where many of the present elderly have minimal or no education. Cognitive tests that are designed to be less susceptible to cultural influence, such as the Rowland Universal Dementia Assessment Scale [1] and the Clock Drawing Test (CDT), are most pertinent. The CDT has shown promise as an ideal cognitive test, as it is quick to administer and appears to be less vulnerable to linguistic, ethnic, cultural or educational bias than the MMSE [2–3]. GPs have indicated that they would be likely to use the CDT as a screening tool if proven ‘effective’ [4]. Intuitively, the CDT appears useful to overcome the language and education barriers in our population. Nonetheless, some have expressed doubts about the usefulness of the CDT, while others have noted that the task can be significantly affected by low education [5–7].

The CDT assesses a wide range of cognitive functions, including memory, language, visuoperceptual and visuconstructional praxis, and importantly, executive functions such as planning, organization, simultaneous processing and self-monitoring [8]. Disturbances in executive control function (ECF) can result in functional decline, with difficulties in instrumental activities of daily living such as cooking, shopping, driving and taking medications [9–10]. Frontal systems and ECF are not well assessed by the MMSE despite these functions being commonly affected even in early Alzheimer’s disease (AD) [11–12]. Moreover, several non-AD dementias such as vascular dementia (VaD) present with ECF deficits early but with less memory deficits. Screening with the MMSE may thus miss these dementias in their initial presentations. Therefore, in 1994, the American Psychiatric Association added ECF to its list of deficits used to establish a diagnosis of dementia [13]. In addition, the CDT has been shown to be able to differentiate between AD and VaD due to its sensitivity to ECF deficits [14]. Given that the usual batteries to assess ECF deficits are usually time-consuming to administer, the CDT can provide a convenient way to detect ECF deficits [15].

There are many variants of CDT, each with its own scoring system [16–20]. There are qualitative differences in how dementia subgroups fail a CDT even if they have similar levels of severity [21]. Some failures arise from true constructional deficits, while others are due more to ECF-related pathology. Most CDT scoring methods do not grade the CDT as an executive task, nor differentiate the executive control of clock drawing from the drawing itself. CLOX may be able to solve these qualitative differences [22]. CLOX is divided into 2 parts. CLOX1 is an unprompted task and CLOX2 a copied version. Apart from its value as a screen for dementia, CLOX provides the capacity to develop 2-dimensional assessments of cognition. The pattern of CLOX1 versus CLOX2 performance may have diagnostic significance not conveyed by either measure alone [22].

We therefore conducted this study with the following aims: (1) to examine the clinical utility of CLOX in detecting dementia in our population and to determine the optimal cut-offs from the receiver operating characteristic (ROC) analysis for CLOX1 and CLOX2, and (2) to analyze known group differences and strength of anticipated associations of CLOX1 and CLOX2 in AD compared to a composite group of AD with cerebrovascular disease (AD with CVD) and VaD. The latter is denoted as vascular composite group (VCG).

Methods

Subjects

The subjects with dementia (n = 73) were referred by primary-care physicians to the Dementia Assessment Clinic of the Department of Geriatric Medicine in the Alexandra Hospital. The community subjects that formed the healthy control group (n = 75) were an aged-matched (≥65 years) subset of the cohort from a community health screening project for elderly residents in the south-east region of Singapore, a component of an ongoing cohort study, the Singapore Longitudinal Aging Study. Residents in the defined survey area were identified in door-to-door census and invited to participate in the health examination. The response rate was estimated at 78.5%. They completed extensive questionnaires and tests that included demographic, health, behaviour, functional and cognitive assessments. Clinical dementia rating (CDR) [23] was applied to the control group to exclude dementia and mild cognitive impairment (MCI).

All subjects were ≥65 years old with normal vision and hearing. Normal vision was defined as being able to read newsprint, and if the patient was able to carry out a normal conversation, he was considered hearing intact. Only patients with mild to moderate dementia defined by Diagnostic and Statistical Manual of Mental Disorders, 3rd edition revised [24] were included. In the healthy controls, exclusion criteria were a score <27 in the MMSE, a CDR score >0, history of significant head injury, stroke or evidence of CVD (Hachinski score >4) [25], other neurological disease, systemic illness or medical conditions that may affect cognitive functioning and activities of daily living, clinical depression (score of ≥12 on Hamilton Depression Rating Scale) [26] or other psychiatric and substance-related disorders which affect cognitive functioning, and use of long-acting benzodiazepines or barbiturates within the past 2 years.

Diagnosis and Assessments

The diagnoses were established according to DSM-IV criteria for dementia [13] by a consultant geriatrician in the Dementia As-

Yap/Ng/Niti/Yeo/Henderson
assessment Clinic on the basis of medical history, clinical examination, relevant blood investigations and brain imaging with either CT scan or MRI. Patients with questionable dementia, mild dementia and dementia that was atypical of AD were further subjected to a battery of neuropsychological assessments by a clinical psychologist. Thereafter, a consensus meeting, comprising a panel of 2 geriatricians, a clinical psychologist and a dementia nurse clinician, determined the final diagnosis of dementia and the type of dementia. The following diagnostic criteria were used to define the subtypes of dementia, DSM-IV and National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer’s Disease and Related Disorders Association for AD and AD with CVD [27], and National Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherche et l’Enseignement en Neurosciences for probable VaD [28]. Hachinski’s ischaemic score was also assessed for all patients with dementia. CLOX was rated by a single rater and the dementia diagnosis was made independently of the CLOX scores to ensure there was no circularity in the diagnostic process.

CLOX is a CDT developed by Royall et al. [15]. We employed the Chinese translation of CLOX [29] for our study. The test is divided into 2 parts, CLOX1, an unprompted clock drawing task, and CLOX2, a copied version. In CLOX1, the subject is asked to draw a clock face on a blank surface at the back of the CLOX instrument when instructed: ‘Draw me a clock that says 1:15. Set the hands and numbers on the face so that a child could read them.’ In CLOX2, the examiner draws the clock face required in a pre-drawn circle and the subject is asked to copy it in an adjacent blank area. CLOX is scored with respect to (1) its resemblance of visual attributes to a standard analogue clock and (2) the constructional sequence (making the 4 quadrants by placing the numbers 12, 6, 3 and 9 first in order to facilitate the positioning of the entire number set). Each CLOX subtest is scored on a 15-point scale. Lower CLOX scores denote greater cognitive impairment.

Data Analysis

The 1-sample Kolmogorov-Smirnov test was performed to determine whether the CLOX scores were normally distributed. Validation of CLOX was done by subjecting CLOX1 and CLOX2 scores of normal controls and patients with dementia to ROC analysis to determine optimal cut-off scores with the corresponding sensitivities and specificities. Stratification by education into 2 groups, 0–6 years and >6 years, was carried out. Further analysis was performed to compare the diagnostic accuracy of CLOX1 and CLOX2 for the 2 subgroups of dementia, namely AD and VCG. Analysis of variance was used to compare CLOX test scores between controls, AD and the VCG.

Results

Subjects’ Characteristics and CLOX Performance

The total sample size was 148, comprising 73 subjects with dementia (49.3%) and 75 healthy controls (50.7%). The demographic characteristics with respect to age, gender and educational status of the 2 groups and their mean MMSE scores are shown in table 1. The dementia group was older and had fewer years of education. Mild- (52.1%) and moderate-stage dementia (47.9%) were fairly equally represented. As regards dementia subtypes, 47 (64.4%) had AD, while 26 (35.6%) comprised the VCG. CLOX scores were normally distributed, as suggested by the statistically significant p values of the Kolmogorov-Smirnov test (p < 0.05 for both CLOX1 and CLOX2). Overall, the mean CLOX2 scores were higher than the CLOX1 scores, as seen in table 2. The CLOX scores were lower in those ≥75 years old, females and those with less education. The Hachinski scores were highest for VaD (8.7, SD = 1.6), intermediate for AD with CVD (4.1, SD = 2.1) and lowest for AD (1.9, SD = 1.6).

Validation of CLOX and CLOX Profiles for AD and VCG

Table 3 shows the diagnostic accuracy of CLOX1 and CLOX2 against clinically diagnosed dementia. The areas under the curve (AUC) for CLOX1 and CLOX2 were 84 and 85%, respectively. The diagnostic accuracy of CLOX was greater in those with >6 years of education compared to those with 0–6 years, as shown by the generally higher AUCs. For CLOX1, the AUC was greater for VCG (88%) compared to pure AD (81%). This difference was minimal

<table>
<thead>
<tr>
<th>Table 1. Demographic characteristics (n = 148)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristic</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Dementia subtype</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Dementia severity</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Figures in parentheses are percentages.
in CLOX2, 86% in VCG compared to 85% in AD. The mean CLOX scores were lower in the VCG compared to AD, and this difference was more pronounced in CLOX1 compared to CLOX2, as shown in Table 4. After adjustment for the confounding variables of age, gender, education, MMSE and dementia stage, the difference in CLOX scores was found to be statistically significant for CLOX1 (p = 0.002) and non-significant for CLOX2 (p = 0.074). Additionally, the magnitude of the difference in mean CLOX scores between CLOX1 and CLOX2 was greater in VCG (3.1) compared to AD (2.2).

Figure 1 shows the ROC curves for CLOX 1 and 2 validations. The optimal cut-off was 10 for CLOX1, yielding a sensitivity of 75% and a specificity of 76%. For CLOX2, cut-off at 12 carried a sensitivity and specificity of 75 and 80%, respectively. The optimal cut-offs for CLOX stratified for education are shown in Table 5. For those with 0–6 years of formal education, CLOX 1 and 2 cut-offs remained at 10 and 12, respectively, yielding sensitivities of 78.6 and 76.8% and specificities of 64.9 and 73.0%. The cut-offs were higher in the group with >6 years of education, being at 11 for CLOX1 and 13 for CLOX2 with sensitivities of 88.2 and 87.5% and specificities of 68.4 and 87.5%.

The likelihood ratio of a positive test (LR+) was 3.1 for CLOX1 and 3.8 for CLOX2. In those with >6 years of education, a cut-off of 13 for CLOX2 yielded the highest LR+ of 4.8. The positive predictive value (PPV) exceeded 75% in both CLOX1 and 2 with dementia prevalence at 49.3% in the sample.

### Table 2. CLOX performance by socio-demographic characteristics

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Mean CLOX1</th>
<th>p value</th>
<th>Mean CLOX2</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>148</td>
<td>9.5 (4.3)</td>
<td>–</td>
<td>11.7 (3.8)</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65–74 years</td>
<td>75</td>
<td>11.2 (3.2)</td>
<td>p &lt; 0.001</td>
<td>13.1 (2.5)</td>
</tr>
<tr>
<td>≥75 years</td>
<td>73</td>
<td>7.7 (4.6)</td>
<td></td>
<td>10.2 (4.4)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>57</td>
<td>10.9 (3.7)</td>
<td>p = 0.001</td>
<td>12.6 (3.2)</td>
</tr>
<tr>
<td>Female</td>
<td>91</td>
<td>8.6 (4.5)</td>
<td></td>
<td>11.1 (4.1)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–6 years</td>
<td>93</td>
<td>8.3 (4.5)</td>
<td>p &lt; 0.001</td>
<td>10.8 (4.1)</td>
</tr>
<tr>
<td>&gt;6 years</td>
<td>55</td>
<td>11.5 (3.3)</td>
<td></td>
<td>13.2 (2.7)</td>
</tr>
</tbody>
</table>

Figures in parentheses represent SD.

### Table 3. Diagnostic performance of CLOX by education

<table>
<thead>
<tr>
<th></th>
<th>All subjects</th>
<th>0–6 years education</th>
<th>&gt;6 years education</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLOX1</td>
<td>0.84 (0.77–0.90)</td>
<td>0.78 (0.69–0.88)</td>
<td>0.88 (0.80–0.97)</td>
</tr>
<tr>
<td>CLOX2</td>
<td>0.85 (0.79–0.91)</td>
<td>0.81 (0.72–0.90)</td>
<td>0.91 (0.84–0.98)</td>
</tr>
</tbody>
</table>

Values are AUC (with 95% CI in parentheses).

![Fig. 1. ROC curves for CLOX1 and CLOX2 against clinically diagnosed dementia.](image-url)
Discussion

The AUCs for CLOX 1 and 2 were good, close to 85% in both instances. The AUC is a better measure of predictive accuracy than sensitivity and specificity, as it yields an index independent of cut-off points and disease prevalence. Thus, CLOX demonstrates adequate diagnostic accuracy for dementia in our sample of elderly with mild to moderate dementia. It is noteworthy that the cut-offs for CLOX1 and CLOX2 in our study correspond to the 5th percentile in the adult reference group of Royall et al. [15]. This supports the notion that CDTs, at least by CLOX, are consistent across cultures and appear little affected by cross-cultural and inter-ethnic differences. The same cannot be said of the MMSE, which, being reliant on language, is clearly influenced by culture. This makes...
CLOX an attractive cognitive screening test in countries that comprise peoples of multiple ethnicities such as ours. Notably, the CLOX performance was poorer in females, in those ≥75 years old and those with less education. Older people are at greater risk of dementia, hence the poorer CLOX scores.

Moreover, the poorer performance in females and older people is likely mediated through education as well. In our population, the literacy rates were low in the cohort born in the early 1900s and women were given less opportunity for formal education compared to men in that era.

The sensitivities for CLOX 1 and 2, both at 75%, are only modest, while CLOX2 shows greater specificity at 80%. Our findings are comparable to those in a previous study [30], which demonstrated a sensitivity of approximately 80% against clinical dementia among 5 different CDTs. Royall et al. [15] found that CLOX subtests had a discriminative power of 83% in differentiating normal and AD elderly and this increased to 92% in discriminating AD subgroups with and without constructional impairments. The specificity of CLOX drops in those with less education, suggesting a greater likelihood of a false-positive test in this group. Although CDTs are believed to be less susceptible to the effects of education, the results show a clear influence of education on CLOX scores in our sample. For this reason, stratifying by education was performed in establishing cut-offs for dementia and CLOX showed satisfactory sensitivities of 87–88% in those with >6 years of education. Literacy still has an effect on CDTs because those who are illiterate may be less familiar with writing and drawing. The CDT is also dependent on good vision and motor skills that may be deficient in the elderly, especially the ‘old-olds’ (age ≥75 years). Our experience in administering CLOX reveals the difficulties faced by some subjects. They tend to hesitate and have problems initiating the task, remarking that they have never held a pen, and need gentle coaxing to proceed. Some encounter problems in writing numbers and their relative inexperience in drawing often shows up as spatial inaccuracies in the clock. Nonetheless, as the demographic patterns indicate higher literacy rates in future cohorts of older persons, this issue will gradually hold less significance.

The quality of the test, reflected by the PPV, both exceeded 75% but was slightly higher in CLOX2 (78.3%) compared to CLOX1 (75.3%) given a dementia prevalence of 50.7%. This is satisfactory, as the probability of dementia in clinical samples of patients seen for memory complaints is expected to be high. The PPV was higher in those with 0–6 years of education mainly because this group was more represented (60.2%) in our sample. Overall, CLOX2 appears to have a greater predictive value in our sample mainly because of its greater specificity compared to CLOX1, with less resultant false-positive values.

The CLOX1 AUC in VCG (88%) was higher compared to pure AD (81%). The difference in AUC was less in CLOX2, 86% in VCG versus 85% in AD. The mean CLOX1 scores were lower in VCG than AD, even after adjustment for the covariates age, gender, education, MMSE and dementia stage (p = 0.001). Although there was likewise a difference in CLOX2 scores between VCG and AD, it was smaller and not statistically significant. The magnitude of difference in mean CLOX 1 and 2 scores was also greater in VCG (3.1) compared to AD (2.2). Thus, on the understanding that there is likely greater ECF impairment in the VCG compared to pure AD, the results appear to support the rationale that CLOX1 (clock drawing to command) is more sensitive to ECF deficits compared to CLOX2 (clock drawing to copy), a principle upheld by Royall et al. [15]. However, others have asserted that in fact the relationship between clock drawing and ECF is found more in the copy than the command condition [31, 32]. Nonetheless, the disparity in CLOX profiles between VCG and AD supports the notion that AD patients as opposed to patients with vascular pathology tend to perform differently in the draw to command compared to the copy task of the CDT and that CLOX is indeed sensitive to ECF in our sample. This is in contrast to the CLOX study in Hong Kong Chinese, where CLOX was not shown to be sensitive to ECF deficits in patients with subcortical ischaemic vascular disease [29]. Recent studies suggest ECF appears to be affected early in AD as well and may be the common feature underlying all dementias [33, 34]. Therefore, the ability of a screening test to assess ECF deficits is useful in detecting cognitive impairment early. It is in this respect that CLOX can fulfill an unmet need that commonly used cognitive screens such as MMSE and Abbreviated Mental Test fail to address [35].

Several limitations of this study should be noted. First, the criterion for MCI was not applied to definitively exclude MCI in the control group, although a CDR score of 0 can be considered an adequate measure. Second, despite the attempt to match controls and dementia subjects for age (≥65 years), the control group was still younger than the dementia group. There was also a significant difference in the education levels between the 2 groups. These factors would have affected the validity of
the results. Third, in the VCG, the numbers of each subtype of dementia are small. It would be ideal if more subjects were available for CLOX performance to be compared across the individual dementia subtypes of AD with CVD and VaD instead of lumping them together. These dementia subtypes are grouped together on the basis that they are likely to have more ECF deficits compared to pure AD, matched for severity. However, no other standard test of ECF has been applied to the subjects to better assess the presence of ECF deficits. Therefore, the conclusion that CLOX is sensitive to ECF in our sample is based on the understanding that ECF deficits are more prominent in AD with CVD and VaD. Finally, it should be remembered that CLOX remains a screening test for dementia and only aids as an adjunct in differentiating diagnostic dementia categories. It is surely not a definitive means to diagnose the dementia subtype in itself.

**Conclusion**

The need for a quick, convenient and reliable test that will encourage cognitive screening in busy GP clinics prompted this study. Furthermore, CDTs are already used by many clinicians in memory clinics, but no objective scoring method has been validated in our country. Our results support CLOX as a valid screen for dementia, not forgetting the particular importance of a reliable informant history, and conditions that can confound CLOX scores, such as depression and delirium, must be excluded. The results also suggest that CLOX has the added advantage of being more sensitive to ECF deficits than other conventional cognitive screening tests. The main limitation appears to be administering the test to those who are illiterate and are less able to draw and write. Education appears to have a significant bearing on the CDT in our present cohort of elderly. Nevertheless, this issue will become less important as the level of education increases in future population cohorts. CLOX holds much promise as a cognitive screen and should be applied on a wider basis in our population.

**References**


