Examination of The Impact of Education on Cognitive Screening Tests

Mairi Crombie, BSc (Hons), MSc, PgDip

Submitted in partial fulfilment of the requirements for the degree of

Doctorate in Clinical Psychology

Institute of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

March 2021

TABLE OF CONTENTS

List of Tables Included

Chapter 1: Systematic review

Chapter 2: Major Research Project

List of Figures Included

Chapter 1: Systematic Review

Chapter 2: Major Research Project

Acknowledgments

Firstly, I would like to express my gratitude to Professor Jon Evans who has been a constant source of support and has gone above and beyond throughout this process. I feel very fortunate to have had someone so passionate, knowledgeable and experienced to supervise me. It was a wonderful opportunity to work alongside his colleagues from Kolkata, India; Dr Aparna Dutt, Ms Priyanka Dey and Ms Ranita Nandi; thank you to you all, I could not have done this project without you.

I would like to thank all the Highland trainees and assistants for their support throughout training.

Finally, I would like to thank my mother, Patricia and my sister, Ailsa who have been there for endless support and encouragement always and throughout this journey.

Foreword

To provide context for this project; circumstances related to the onset of Covid-19 led to an interruption of my original MRP resulting in me not being able to recruit participants (See appendix 2.6 and 2.7 for original MRP proposal), therefore this project was abandoned.

In line with guidance from the University of Glasgow, I carried out a study involving a new analysis on an existing data set. The data were originally collected by research colleagues of my supervisor Professor Jon Evans (Dr Aparna Dutt, Ms Ranita Nandi and Ms Priyanka Dey) in Kolkata, India.

CHAPTER 1: THE IMPACT OF EDUCATION ON THE CLOCK DRAWING TEST (CDT): A SYSTEMATIC REVIEW

Mairi Crombie*

BSc, MSc, PgDip

***Address for Correspondence**

Institute of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

1st Floor, Administration Building,

Gartnavel Royal Hospital

1055 Great Western Road

Glasgow, G12 0XH

Tel: 0141 211 0607

Email:

Word Count (including tables, figures, references):7425

Prepared in accordance with guideline for submission to "The Clinical Neuropsychologist" (See appendix 1)

ABSTRACT

Background and Objectives

The Clock Drawing Test (CDT) is often utilised in the screening and assessment of cognitive impairment and dementia. Issues arise as it is used in countries where variation in education level is greater than where it was originally developed and validated, highlighting the importance of considering evidence that education impacts CDT performance. This systematic review therefore examined the impact of education on CDT performance.

Method

Studies were identified through searching MEDLINE, EMBASE, PsycINFO, PubMed, Web of Science, CINAHL, Cochrane and Scopus. Inclusion and exclusion criteria were applied. Risk of bias was assessed using an adapted version of the Quality Assessment Tool for Systematic Observational Studies (QATSO; Wong et al., 2008).

Results

Twenty-one papers were identified. Twenty (95%) studies found a relationship between education and performance on the CDT. Most studies were of acceptable methodological quality. Effect sizes varied considerably but were mainly medium to large in size.

Conclusion

There is good evidence that CDT performance is affected by education, specifically amongst those with limited education. It would therefore be useful to have an alternative to the CDT that is sensitive to similar cognitive domains without requiring minimum levels of education or literacy skills.

Keywords

Clock drawing test, CDT, education, illiterate, literate

INTRODUCTION

Life expectancy is increasing, creating a larger population of older adults throughout the world; however, the pattern and pace of this growth varies in different countries. It is predicted that significant growth in aging populations will occur in low to middle income countries (LMIC), such as India, in the coming years. Growth will be slower in the countries of the West, that have already seen an increase (United Nations Department of Economic and Social Affairs, 2017).

An aging population increases the risk of age-related diseases such as dementia (Alzheimer's Disease International, 2010). Identifying dementia or cognitive difficulties associated with aging at an early stage is important as it maximises opportunities for intervention and preparation for later difficulties. It is therefore essential that we have effective screening and assessment tools to identify cognitive difficulties.

An ideal cognitive impairment screening test should be relatively independent of culture, language, and education, whilst being quick to administer, easy to score and well tolerated and acceptable to patients (Shulman, 2000). Traditionally, the Clock Drawing Test (CDT) has been used to assess the mental status of patients for various neurological difficulties (Freedman et al., 1994) and has long been used as a screening method for cognitive impairment (Shulman, 2000), especially in the elderly (Nishiwaki et al., 2004). It is a brief tool that has been applied internationally, often used in the assessment of individuals suspected of having dementia (Royall et al., 1998). Although there are multiple versions of this test, in general, they all require the patient to draw the face of a clock and then to draw the hands to indicate a particular time. This single test may be sensitive to dementia because it involves many cognitive areas that can be affected by dementia, including executive function, visuospatial abilities, motor programming, attention and concentration. The CDT is included as a subtest in cognitive screening tests such as the Addenbrooke's Cognitive Examination (ACE III; Hsieh et al., 2013).

Evidence indicates that CDT results may be influenced by education level as performance requires familiarity with using a pen or pencil along with writing numbers (Kim and Chey, 2010; Nielsen & Jorgensen, 2013) which can lower the specificity of the test. Various studies (e.g., Tripathi et al., 2014) have demonstrated the influence level of education has on performances of neuropsychological tests, meaning that tests may not be valid for those with low or no education or who are illiterate.

Some researchers (e.g., Shulman et al., 1986) have stated that performance on the CDT is free from educational and cultural bias. Several other studies, however, have found that performance on the CDT may be influenced by level or years of education, particularly in people with little or no education or those who are illiterate. Borson et al. (1999) found a significant influence of the level of education on the performance of the CDT.

Many scoring systems of the CDT exist. A literature review including scoring systems of the CDT by Pinto & Peters (2009) included only three papers examining the effects of level of education on CDT

performance, and all displayed a significant relationship. Brodaty & Moore (1997) found a correlation of the CDT score with years of education when using two versions, Shulman (1993) and Sunderland (1989) scoring systems, but not the Wolf-Klein (1989) scoring system. Two of these studies (Brodaty & Moore, 1997; Borson et al. 1999) involved grouping patients and healthy controls together when looking at the influence of education. For this review, we will focus on studies that included analyses of healthy control samples only or separately from patient groups.

The present systematic review therefore aimed to identify and synthesise the extant literature examining the relationship between education and performance on the CDT. There are no recent systematic reviews of this literature in which the strength of the association between education and CDT performance is systematically synthesised.

METHODS

Papers were identified from a search on the following databases: MEDLINE, EMBASE, PsycINFO, PubMed, Web of Science, CINAHL, Cochrane, Scopus (see Figure 1) on 22nd November 2020. Duplicates were removed. The title and abstracts of all the papers were screened for eligibility based on the inclusion and exclusion criteria. The reference lists of all included studies were also searched for additional relevant studies. These studies were also used to identify additional studies using the related article feature of databases. The review was registered with Prospero (CRD42020222113) and PRISMA Guidelines (Moher et al., 2015) were followed (see Figure 1).

Inclusion criteria

We sought to identify papers that investigated the variable education along with performance on the CDT. Only studies that compared healthy controls and education on the CDT were included for final review.

Exclusion criteria

The article search was limited to papers published in English. Studies that investigated a patient group (e.g., dementia patients) alone were excluded. Papers which failed to report the normative data or the relationship between CDT performance and education were also excluded.

Databases MEDLINE, EMBASE, PubMed, PsycINFO, Web of Science, CINAHL, Cochrane, Scopus were searched using the following terms:

(i)“educat*” OR “literat*” OR “illiterat*” OR “educat* status” OR “status educat*” AND (ii) “Clock draw*” OR “clockdraw*” OR “clock complet*”

Methodological quality appraisal

A quality assessment tool (see Appendix 1.1.1) was designed for this review, based on the Quality Assessment Tool for Systematic Observational studies (QATSO; Wong et al., 2008). This tool aimed to assess the risk of bias for the following domains; design of the study, confounding variables accounted for, validity of predictor measure and validity of outcome measure. Adaptations were required due to gaps in existing tools that failed to consider all variables deemed necessary to gain a thorough assessment of quality.

Quality assessment scores were calculated, and higher totals reflected higher quality studies, with a maximum score of 8 (see Appendix 1.1.2). Studies achieving 67% or more in the score were regarded as "good" quality; 34–66% "fair"; and, below 33% were regarded as "poor" (as per Wong et al.'s (2008) guidelines). Studies included for the final review were scored for methodological quality by the first author and 33.33% (n=7) were scored by a peer reviewer (see Appendix 1.1.3). Differences in ratings did not exceed one point in all studies rated. Any discrepancies were discussed and resolved.

RESULTS

Figure 1 provides an overview of the search, screen and eligibility assessment process followed within this review (following PRISMA guidelines). Studies that involved an assessment of education as well as performance of the CDT were included. The primary outcome was the association between CDT performance and level of education (in years) and literacy status.

A total of twenty-one studies were included for the final review based on the inclusion criteria i.e., studies which included a sample of healthy controls and a comparison of CDT performance against the variable education. Table 1 provides a summary of study details and results for all included studies.

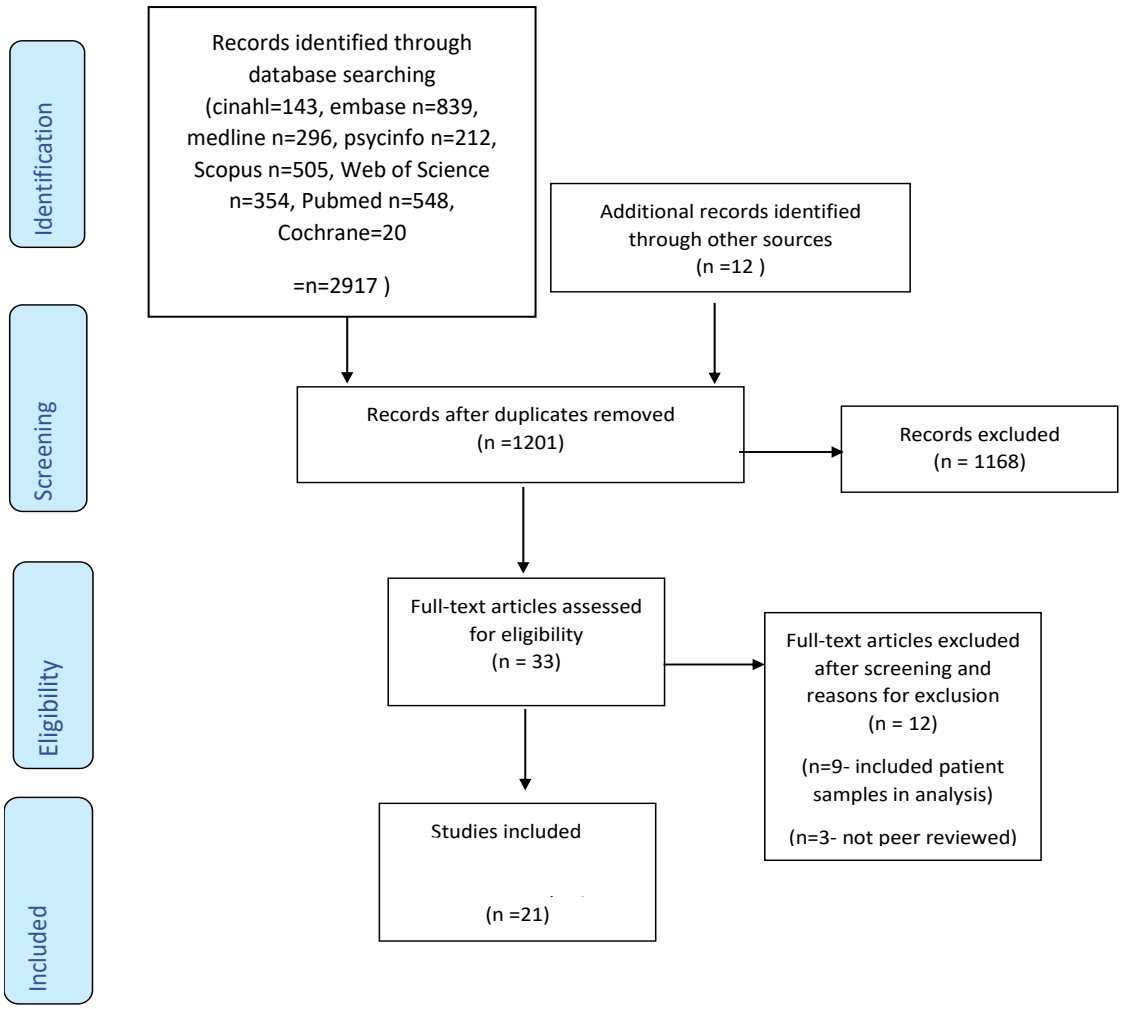


Figure 1: Prisma flow diagram of search strategy

Out of twenty-one studies, twenty found a relationship between education and CDT performance. One of these studies (Sugawara et al., 2010) had five different conditions of drawing the clock, only one of these conditions was shown to be influenced by level of education.

Across the studies there was a wide range of effect sizes evident, from small to large (see table 1). Correlation coefficients (r) were reported for some studies. Standardized regression coefficients (beta values) derived from multiple regression analyses were reported in some of the studies. Standardized mean differences (Cohen's d) were calculated for studies comparing groups of participants with different levels of education where possible. Correlation effect sizes (r) values were calculated where beta was reported, as recommended by Peterson & Brown (2005).

Some studies found a small effect size (Turcotte et al., 2018; Shanhu et al., 2019) with Sugawara et al. (2010) and Paganini-Hill et al. (2001) reporting particularly small effect sizes. A small to medium effect

size was reported by Von Gunten et al. (2008) and Hubbard et al. (2008). Medium effect sizes were seen in Santana et al. (2013) and Seigerschmidt et al. (2002). Medium to larger effect sizes were seen in Bozikas et al. (2008)'s five conditions and in Leung et al. (2005). Large effect sizes were seen in Nitrini et al. (2004) (when comparing an illiterate with a literate sample), Fabricio et al. (2013) (when comparing the lowest educated with the highest educated), De Noronha et al. (2018), Kim & Chey (2010), Marcopolus et al. (1997), Merims et al. (2018), Sicialano et al. (2016) and in Balduino et al. (2020) (when comparing lowest educated with highest educated).

One paper (Ainslie & Murden, 1993) reported a medium effect size for two scoring systems of CDT (a modified version of Shulman et al. (1993) and scoring system devised by Sunderland et al. (1989)) whilst this study showed a large effect size when the original Shulman et al. (1993) scoring system was used. This study further found that the Wolf Klein (1983) scoring system was least educationally impacted. Shao et al. (2020) displayed a small effect size for the MoCa scoring system, whilst displaying medium effect sizes when using Rouleau's and Babin's scorings systems. Hubbard et al. (2008) displayed a small to medium effect size when analyzing the influence of education whilst using the Cahn Global (1996) scoring system. However, this study did not find an influence of education on Mendez et al. (1992) or Freund et al. (2005) scoring systems. Caffara et al. (2001) failed to provide the required statistical values to allow calculation of an effect size.

Table 1: Summary of studies included in review including sample demographics, methods, measures used and results

Author, year & Location	**Sample demographics (N-HC; N-female, male; age; education level)**	**Type of sample**	**Measurement of education**	**CDT scoring system(s)**	**Results**
Ainslie & Murden, (1993) USA	N=110 Gender and age not reported Education: 79 (72%) LI (8 years or less 31 HI (9+)	Non-demented elderly participants from three university medical centre geriatric divisions.	Two levels: 1.9+ years of education, 2.8 or fewer years of education	Shulman et al.(1993) Modified version of Shulman et al. (1993) Sunderland et al.(1989) Wolf-Klein (1989)	Clock-drawing ability is affected by education in non-demented elderly persons. The scoring method of Wolf-Klein is least educationally affected. Comparison of well and poorly educated non-demented by chi-square shows significant differences on Shulman and Sunderland, but not on the Wolf-Klein scale: Shulman standard, ChiSq= 15.9, $p<0.001$- Cohen's d (computed) = 0.822 Shulman modified, ChiSq =8.7, $p<0.003$ – Cohen's d (computed)=0.58 Sunderland, ChiSq =7.6, $p<0.006$- Cohen's d (computed)=0.54 Wolf-Klein (Statistical value not given) Fisher's exact test, n.s.

Author, year & Location	**Sample demographics (N-HC; N-female, male; age; education level)**	**Type of sample**	**Measurement of education**	**CDT scoring system(s)**	**Results**
Balduino et al. (2020) Brazil	N=144 (gender not reported) Age:80+ years, max 103 Education:range <1->5 years	Patients from the Geriatric Division of the Jundiaí Medical School	Three levels: 1. ILLITR (<1 year of school) 2. 1-4 years 3. >5 years	Mendez et al. (1992)	There was a positive influence of educational level on the CDT scores- Kruskal Wallis P=<.001 Cohen's d- ILLITR vs 1 to 4- d = 1.559 ILLITR vs 5+- d=1.667 1-4 vs 5+ - d=0.369

Author, year & Location	Sample demographics (N-HC; N-female, male; age; education level)	Type of sample	Measurement of education	CDT scoring system(s)	Results
Bozikas et al. (2008) Greece	N=223 (F=110,M=113) Age: (M = 45.99;SD = 18.82; range 17–80) Education: range: 1-13+ years	Healthy community-dwelling adults (volunteers)	Three levels: 1. 1-9 years 2. 10-12 years 3. 13+(uni) years	Freedman et al. (1994). Including 5 conditions *	Regression displayed that education had an influence on CDT performance for each of five forms of CDT (r scores calculated) Clock A β =0.299 p<0.001; r=0.349; Clock B β =0.326 p<0.001; r=0.379; Clock C β =0.257 p<0.00; r=0.307; Clock D β =0.448 p<0.00; r=0.498; Clock E β =0.356 p<0.001; r=0.406
Caffara et al., (2001) Italy	N= 248 (F=124, M=124) Age=20-89 years Education: range =5-13+	Young participants -students and employees from university and older volunteers from church and senior citizen group.	Four levels: 1. 0-5 years 2. 6-8 years 3. 9-13 years 4. 13+ years	Freedman et al. (1994)	No influence of education on CDT performance No stats values reported

Author, year & Location	Sample demographics (N-HC; N-female, male; age; education level)	Type of sample	Measurement of education	CDT scoring system(s)	Results
De Norohna et al. (2018) Brazil	N= 121 (F= 63, M=58) Age (M:39.6 years ;SD = 11.9; range 19-59) Education: range: 1->11 27illiterate adults with 0 years of formal education 34 adults with 1-4 years 30 adults with 5-11 years; and 30 adults with >11 years of formal education	Healthy volunteers selected from individuals accompanying patients of the Acquired Speech and Language Neurological Disturbances outpatient unit and other clinics within the Department of Speech, Language and Hearing Sciences	Four levels: 1.illiterate adults, 2.1-4 years 3. 5-11 years 4. >11 years formal education.	Sunderland et al. (1989)	Significant differences on the CDT were found only between the illiterate and other educated groups- Kruskal Wallis= $p<0.001$ Effect sizes; Illiterate vs 1-4 – d=0.88 Illiterate vs 5-11- d=0.87 Illiterate vs 11+ - d=1.367

Author, year & Location	Sample demographics (N-HC; N-female, male; age; education level)	Type of sample	Measurement of education	CDT scoring system(s)	Results
Fabricio et al.(2013) Brazil	N=141 (F=109, M=32) Age: (M=65.18, SD=6.68) Education (M= 9.38; SD 4.5; range=1-25)	Recruited from memory workshop of wide range of age and education in university study recruited from 2007-2010	Three levels: 1.1-4 years 2. 5-8 years 3. 8+years	Shulman et al. (1993) Sunderland et al. (1989) Rouleau et al. (1992; Modified version)	Kruskal Wallis test comparison of 3 CDT measures with 3 educational levels: p=0.008 (Shulman) 1-4 vs 5-8=Cohen's d=0.68; 1-4 vs 8+=Cohen's d=0.76 5-8 vs 8+=Cohen's d= 0.01 p=0.002 (Sunderland) 1-4 vs 5-8=Cohen's d=0.65 1-4 vs 8+=Cohen's d=0.88 5-8 vs 8+=Cohen's d= 0.172 p=0.004 (Rouleau) 1-4 vs 5-8=Cohen's d=0.615; 1-4 vs 8+=Cohen's d=0.825 5-8 vs 8+=Cohen's d= 0.136

Author, year & Location	Sample demographics (N-HC; N-female, male; age; education level)	Type of sample	Measurement of education	CDT scoring system(s)	Results
Hubbard et al. (2008) USA	N=207 (F=134,M= 73) Age: (M=71.3; S.D. 8.4) Education: (M=16.6, S.D.=2.7; range=11-24 years)	Healthy volunteers enrolled in the patient control registry for the Boston University Alzheimer's Disease Core Centre	Two levels: 1. Low education (no degree) 2. High education (degree +)	Mendez et al. (1992) Freund et al. (2005) Cahn et al. (1996)	Education level was not significantly related to CDT scores using the Freund or Mendez systems, but did relate to scores using the Cahn system Independent samples t-test value (low education vs high education) for Cahn global t (205) =-2.48 p=0.014 Cohen's d (calculated)=-0.38

Author, year & Location	Sample demographics (N-HC; N-female, male; age; education level)	Type of sample	Measurement of education	CDT scoring system(s)	Results
Kim & Chey (2010) Korea	N=240 (F=166, M=74) Age: (M=69.13; SD =8.11) Education: (M=7.48; S.D=5.06); range=1-12)	Elderly Korean people with a range of educational levels (volunteers)	Two levels: 1.Years of formal education ≤ 6 2. ≥ 7	CSS modified – Todd et al., (1995) Rouleau et al. (1992) (qualitative scoring only)	Educational attainment and literacy status of older people influenced performance on the CDT significantly (p < .001) Literacy F = 27.17 p<.001 Regression β=0.23 p<.001 Education F = 9.64 p<.001 Regression β =-.47 p<.001 Illiterate (1-6 years) vs 7+ years (literate) -d = 2.36 Illiterate (0 years) vs 7+ years (literate) – d=2.16 Illiterate (1-6 years) vs 1-6 years (literate) – d=1.61 Illiterate (0 years) vs 1-6 years (literate)- d=1.08

Author, year & Location	Sample demographics (N-HC; N-female, male; age; education level)	Type of sample	Measurement of education	CDT scoring system(s)	Results
Leung et al. (2005) Hong Kong	N=66 (M & F not reported) Age : (M=74.9; SD=4.64; range -68-87 years) Education: (M=4.47, S.D=4.97; range-less than 2- more than 6 years	Community dwelling elderly, recruited from elderly social centre	Three levels 1.=>2 years 2.=2-6 years 3. Middle school or higher – more than 6 years	Chinese Clock Drawing (Lam et al (1998) (correlated scores with Shulman et al. (1993)	Correlation between education and CDT; clock drawing r = -0.53; p<0.01 clock copying r=– 0.44, p<0.01
Marcopulos et al (1997) USA	NC=133 (F=103, M=30) Age (55+) Education: (M=6.65; S.D=2.14; range=0-10 years)	Recruited from senior centres (approximately 70% of the sample), as well as community centres, homes for adults and retirement communities White and African American, non demented, healthy rural community elders.	Four levels: 1.0-5 years 2. 5-6 years 3. 7-8 years 4. 9-10 years	Libon et al. (1996) Sunderland et al (1989)	Multiple regression analysis found education F=10.69 p=<.001 β=0.25 was a significant predictor of clock drawing r2= 0.25

Author, year & Location	Sample demographics (N-HC; N-female, male; age; education level)	Type of sample	Measurement of education	CDT scoring system(s)	Results
Merims et al. (2018) Israel	N=295 (F=147,M= 133) Age (M= 52.6; SD19.6; range, 20–86 years) Education (M=11.4; SD 4.3; range= 0–20 years)	Community dwellers Israeli Arabs- healthy adults and healthy elderly	Years of education	Freedman et al. (1994)	Positive correlation displayed more years of education related to higher scores on CDT (r=0.51-0.62 p<0.01)
Nitrini et al.(2004) Brazil	N=51 (F=27, M=24) Age (M=73.78; SD=5.44) Education: (M=3.82; S.D 3.31; range 1-13 years)	Healthy elderly participants recruited from population-based study that we have been performing in the city of Catanduva, São Paulo State, Brazil (Herrera et al., 2002).	Two levels: 1. Illiterate 2. Literate (subdivided into low educated literate and standard educated literate)	Sunderland et al. (1989)	Comparison of literate vs illiterate (Mann Whitney test) p=.0001- Cohen's d (calculated)=2.46 Illiterate vs low educated vs standard educated (Kruskal Wallis) p=.0001 Low educated literate vs standard educated literate: Cohen's d (calculated)=0.37

Author, year & Location	Sample demographics (N-HC; N-female, male; age; education level)	Type of sample	Measurement of education	CDT scoring system(s)	Results
Paganini-Hill et al. (2001) USA	N=4843 (F=3251, M=1592) Age 80 years (range 52–101 years) Education: 92% completed high school, 37% college graduates	Retirement home – white, well educated, upper middle class from "Leisure World Cohort Study"	Measurement of education included in analysis unclear-possibly a)2 levels: 1.College graduate 2.others (high school completers and non high school completers) OR b) 3 levels 1. college graduates 2. high school 3. non high school completers	Freedman et al. (1994) Watson et al. (1993)	Multiple regression analysis indicated a significant effect of education on CDT) β =0.05; p=.0001 (Calculated) r=0.10

Author, year & Location	Sample demographics (N-HC; N-female, male; age; education level)	Type of sample	Measurement of education	CDT scoring system(s)	Results
Santana et al. (2013) Portugal	N=630 (F=401, M=229) Age (M=55.96; range 25–91) Education: (M=8.08; S.D.=4.58); range=1-12+)	Healthy community-based sample living across Portugal, recruited at the local primary healthcare services and at daycare centres by indication of their general physician. Smaller percentage of subjects volunteered themselves.	Four levels: 1. 1-4 years 2. 5-9 years 3. 10-12 years 4. 12 + years	Rouleau et al. (1992) Cahn et al. (1996) Babin's et al. (2008)	Significant impact of education on results-correlations: Rouleau r=(630)=.405, p<0.001) Cahn r=(630)=.421, p<0.001) Babin's r= (630)=.463, p<0.001).
Seigerschmidt et al. (2002) Munich, Germany	N=139 (gender not reported) Age (M=75; S.D. 5.6: range 65-85) Education: range 9 or less-13 or more years.	Healthy sample recruited from three hospitals in Germany	Two levels: 1. < 9 years 2. > 9 years	Manos &Wu (1994) Watson et al. (1993) Wolf Klein et al. (1983) Shulman et al (1993)	CDT scores were influenced by education: L1 vs L2 (ANCOVA) F=5.07, p<0.01 Male=Cohen's d=0.77 Female= Cohen's d=0.51

Author, year & Location	Sample demographics (N-HC; N-female, male; age; education level)	Type of sample	Measurement of education	CDT scoring system(s)	Results
Shanhu et al. (2019) China	N=885 (M=440, F=445) Age: (Range 65 - 93) Education: range 0-13+ years High-school level group (53.33%), Primary group (24.64%), and the University group (22.03%)	Recruited from communities across 12 counties using stratified random cluster-sampling.	Three levels: 1.Primary – (0-6 years) 2.High school (7-12 years) 3.Uni (13 or more years)	Shulman et al. (1993)'s modified scoring system	Kendall's nonparametric correlation (tau-b) was used to analyse the association of educational level with age and CDT scores (x2 =6.94, p =0.03) Cohen's d (calculated)= 0.18
Shao et al. (2020) China	N=418 (F=165, M=253) Age: (M=63.03; S.D. 7.79; range:35-84) Education: (M=9.32; S.D.=3.14; range=0-18)	Healthy Convenience sample from two residential districts	Two levels: 1.no formal schooling or basic compulsory education 2. ≤9 years, high school 3.10–12 years 4. any university level education, ≥13 years.	MoCA (Kim et al.,2018) Rouleau et al. (1992) Babin's et al. (2008)	In all three scoring systems, CDT scores were significantly correlated years of education MoCa -Kim et al. Correlation (r(417)=.164, p=.001) Rouleau-Regression =((β=0.226, t=5.531, p<0.001) (calculated) r=0.276 Babin's= ((β=0.276, t=5.62, p<.001) (calculated) r=0.326

Author, year & Location	Sample demographics (N-HC; N-female, male; age; education level)	Type of sample	Measurement of education	CDT scoring system(s)	Results
Sicilano et al. (2016) Italy	N=872 (F=483, M=389) Age: range 20–94 years Education: (M=11.17; S.D.=4.98; range=1-27 years)	Healthy participants recruited from different regions; rural, sub-rural, urban etc. - found by advertisements, work, sports centres, educational centres.	Five levels: 1-3 years 4-5 years 6-8 years 9-13 years >13 years	Rouleau et al. (1992)	Regression analyses- F(1,870 = 95.792, P<.001) Cohen's d calculated-(results divided by gender) Male: 1-3 years vs >13- d=1.046 Female: 1-3 years vs >13-d=1.033
Sugawara et al. (2010) Japan	N=873 (F=552, M=321) Age: (M= 57.5; S.D.=11.9; range=30–79 years) Education: (M=11.3; S.D=2.1; range=1-13+)	Participants recruited from 'Health Promotion Project' in 2008.	Three levels: 1.Compulsory education 1-9 years 2. High school 10-12 years 3. University 13+ years	Freedman et al. (1994) * (5 conditions)	Multiple regression analysis displayed that the years of education affected the CDT in the examiner 2 condition:Participants were given three sheets of paper with circles containing the numbers 1–12 and asked to set the hands to 8:20 (:Free drawn-β=0.078, p=0.216; Pre-drawn- β=0.058, p=0.137; Examiner 1- β=0.009, p=0.506; Examiner 2- β=0.067; p=0.001; **r2=0.029= Cohen's d=0.11** examiner 3- β=0.011, p=0.387

Author, year & Location	Sample demographics (N-HC; N-female, male; age; education level)	Type of sample	Measurement of education	CDT scoring system(s)	Results
Turcotte et al. (2018) Canada	N=593 (F=391, M=202) Age: (M = 69.8 years; SD = 7.5; range=43-93), Education:(M = 14.4 years; SD = 3.5; range=5-23)	Healthy community dwelling volunteers in both Montreal and Quebec city	Six levels: 1.Elementary (5–7 years) 2.High school (8–12 years) 3.College (13–14 years) 4.University undergraduate (15–17 years) 5.University graduate (18–19 years) 6. University postgraduate (20–23 years)	Rouleau et al. (1992)	CDT scores significantly correlated with years of education (r(592) = .116, p = .005)

Author, year & Location	Sample demographics (N-HC; N-female, male; age; education level)	Type of sample	Measurement of education	CDT scoring system(s)	Results
Von Gunten et al. (2008) Switzerland	N= 242 (F=175, M=67) Age: (M= 73.4 (SD =8.4) Education: 63 (26.03%) had 'low' levels of education, 86 (35.54%) had 'intermediate' levels of education, and 93 (38.43%) had 'high' levels	Healthy subjects from French speaking region of Switzerland recruited to memory clinic through adverts and referrals.	Three levels: 1.Low 2.Intermediate 3.High	Modified tool based on Montani et al. (1997) Rouleau et al. (1992)	Regression- education on CDT performance $\beta = 0.205$, $p = 0.001$ (Calculated) r=0.225

*Freedman scoring system: A)"Free-drawn"—participants are asked to draw a clock on a blank sheet of paper, fill in the numbers of the clock face and set the hands to 6:45 (Clock A); B)"predrawn"—participants are provided with a predrawn circle and are asked to fill in the numbers of the clock face and set the hands to 6:05 (Clock B); and Clock C – E) Examiner conditions: numbers of clockface are provided, and participants are asked only to set the hands to 11:10 (Clock C) 8:20 (Clock D) and 3:00 (Clock E).

Table 1. Abbreviation Key

M	Male
F	Female
HC	Healthy controls
CDT	Clock Drawing Test

Quality appraisal

The potential risk of bias in each included study was assessed by an adapted version of the Quality Assessment Tool for Systematic Observational studies (see Appendix 1.1.1) (QATSO; Wong et al., 2008). Table 2 provides the QATSO rating percentage for all studies. Wong et al. (2008) set arbitrary cut offs, which we used in this study also; studies achieving 67% or more were regarded as "good" quality; 34–66% "fair" quality; and below 33% as "poor" quality. Seventeen studies therefore were regarded as "good studies". Nine of studies scored above 87.5% (Fabricio et al., 2013; Kim & Chey, 2010; Merims et al., 2018; Santana et al., 2013; Seigershmidt et al., 2002; Shanhu et al., 2019; Shao et al., 2020; Sicialano et al., 2016; Turcotte et al., 2018). One study had a relatively low score of 50% (Balduino et al., 2020), however this was still regarded as "fair" quality, following Wong et al.'s (2008) cut-offs. There did not appear to be any association between quality score and study results. The results from the risk of bias analysis can be found in Appendix 1.1.2. A narrative synthesis of the results is provided below.

Design

Regarding the design of the studies, the majority of studies reported samples of participants from a broad educational range. However, Paganini-Hill et al. (2001); Marcopulos et al. (1997) and Balduino et al. (2020)'s samples were limited and not representative of the population with regard to range of education, e.g., limited to a white, middle class sample.

Confounding variables

Attempts to control for potentially confounding variables (e.g., age, gender) during recruitment or analysis were reported in all studies, except for Ainslie & Murden (1993). De Noronha et al. (2018) and Leung et al. (2005) only attempted to control for one variable. The rest of the studies attempted to control for several confounding variables during recruitment and/or analysis.

Validity of predictor measure

All scores were reduced on validity of predictor measure as all studies failed to include a measure of quality of education in addition to measures of level/years of education.

Validity of outcome measure

All studies used at least one validated CDT scoring system; however, many studies did not use multiple raters to check for reliability of scoring, (Balduino et al., 2020; Bozikas et al., 2008; DeNoronha et al., 2018; Caffara et al., 2011; Marcopulos et al., 1997; Nitrini et al., 2004; Shanhu et al., 2019; Sugawara et al., 2010 and Von Gunten et al., 2008).

Table 2: Ratings of all studies using QATSO tool

Ainslie & Murden (1993)	Balduino et al. (2020)	Bozikas et al. (2008)	Caffara et al. (2011)	De Noronha et al. (2018)	Fabricio et al. (2013)	Hubbard et al.(2007)	Kim et al. (2010)	Nitrini et al. (2004)	Leung et al. (2005)	Paganini-Hill et al. (2001)	Marcopulos et al. (1997)
5/8	4/8	6/8	6/8	5/8	7/8	7/8	7/8	6/8	6/8	6/8	5/8
62.5%	50%	75%	75%	62.5%	87.5%	87.5%	87.5%	75%	75%	75%	62.5%

Merims et al. (2018)	Santana et al. (2013)	Seigerschmidt et al. (2002)	Shanhu et al. (2019)	Shao et al. (2020)	Sicialano et al. (2016)	Sugawara et al. (2010)	Turcotte et al. (2018)	Von Gunten at al. (2008)
7/8	7/8	7/8	6/8	7/8	7/8	6/8	7/8	6/8
87.5%	87.5%	87.5%	75%	87.5%	87.5%	75%	87.5%	75%

DISCUSSION

This review included all studies that have examined performance of healthy individuals on the Clock Drawing Test (CDT) and its association with level of education. In terms of the overall findings, most of the included studies rated high in terms of quality on the QATSO. The QATSO quality results were not associated with the results of the studies. The studies varied widely regarding the scoring methods used; sixteen scoring methods were used across studies in this review, including Rouleau et al. (1992), Mendez et al. (1992), Cahn et al. (1996), and Freedman et al. (1994), Shulman et al. (1993), Sunderland et al. (1989), Manos & Wu (1994), Libon et al. (1996), Moca, (Kim et al. 2018), Wolf-Klein (1989), Freund et al. (2005), CSS modified- Todd (1995), Chinese Clock Drawing -Lam et al. (1998), Babin's et al. (2008), Watson et al. (1993) and a tool based on Montani et al. (1997). The variability of scoring methods means that the results of various studies are difficult to compare (Shulman, 2000). Some studies included more than one CDT rating scale in their analysis (Ainslie & Murden, 1993; Fabricio et al., 2013; Hubbard et al., 2008; Kim &Chey, Marcopulos et al. ,1997; Paganini-Hill et al., 2001 and Santana et al., 2013). A full review of the different scoring systems for CDT is beyond the scope of this review, however, Peters & Pinto (2009) provide a comprehensive review on this topic.

All the included studies apart from one (Caffara et al., 2001) reported an effect of education on CDT performance. In Caffara et al.'s study, the sample were all relatively well educated, and this homogeneity could explain the lack of effect; only 33 out of 248 participants had five years or less of education.

The size of the effect of education on CDT performance varied considerably across studies. Most studies reported medium to large effect sizes. Four of the studies that reported small effect sizes (Paganini -Hill et al., 2001; Shanhu et al., 2019; Turcotte et al., 2018 and Sugawara et al., 2010) reported a relatively high level of education overall. This suggests that even though they may have recruited participants from a broad educational range, this may not have included many participants at the lower end of the range. It is also of note that Sugawara et al. (2010) included five different conditions of the CDT task (using the Freedman scoring system), but only one showed that education had an influence on performance. They found that the effect of education was only obvious in the group of females in the free-drawn circle condition (Sugawara et al., 2014). This study was carried out in Japan where the adult literacy rate is high; most participants (96.8%) had received education for 9 years or more. It is possible therefore that the high level of literacy may have limited the impact of educational difference in CDT score. Another example is Paganini-Hill's (2010) study where there is a particularly small effect size, with most of the participants being from a well-educated, white, middle class sample which included mostly high school or college graduates. Larger effect sizes were apparent in studies with a larger range of education level including participants from the low levels of education, for example, Bozikas et al. (2008), Fabricio et al. (2013), Sicialano et al. (2016) and Nitrini et al. (2004). Nitrini et al. (2004) found

a particularly large effect size, presumably as they compared an illiterate group with a literate group. Balduino et al. (2020) also reported a large effect size, however, it should be noted that the sample in this study are deemed "Super Agers", the authors suggested that this population have a superior memory for their age, including larger cortical volumes and superior resistance to age-related cortical atrophy when compared to people with an "average" cognitive performance of the same age; therefore, these results may not be overly generalisable to the wider population.

There was a lack of eligible papers carried out in LMICs, which limits the conclusions that can be drawn about the use of the CDT as a routine dementia screening tool across cultures and contexts. Four of the studies that reported a large effect size were carried out in Brazil, an Upper-Middle Income Country: Balduino et al. (2020), Fabricio et al. (2013); Nitrini et al. (2004) and De Noronha et al. (2018). This is of note as all the other studies were carried out in High Income countries. We can see a difference of literacy rates in Brazil (at 92% in 2018 compared with 99% in European countries; UNESCO, 2017) therefore, the inclusion of a wider range of education and literacy levels could have influenced the effect size. Interestingly, three out of four of these studies utilised Sunderland et al.'s (1989) scoring system. Authors have previously found Sunderland et al.'s to be one of the more accurate CDT scoring systems (Shulman, 2000).

Variation in effect sizes could also be attributable to the stratification of educational level - most studies have classified education into two levels or more (e.g., Bozikas et al., 2008; Caffarra et al., 2011; Santana et al., 2013 and Siciliano et al., 2016), whereas one study included education as a continuous variable (Merims et al., 2018). The number of categories of educational level varied considerably between studies, with some studies including six levels, (e.g., Turcotte et al., 2018) and others including fewer levels, such as Fabricio et al. (2013), who included three. However, some studies e.g., Turcotte et al. (2018) did not report whether any of their participants were illiterate and started their level of measurement from the minimum five years of education - again this study found a small effect size. Hubbard et al. (2008) defined their 'low education' group as having no degree, and so there was a wide range of educational levels within this category and may not be a robust measurement of low education.

In some studies, the relationship with education was not linear, e.g., Balduino et al. (2020), De Noronha's (2018), Fabricio et al. (2013) and Nitrini et al. (2004). In these studies, the association between education and CDT is linear at the very low levels of education up to the levels of education that are minimal levels in most western countries, but then the relationship with education plateaus as levels of education increase. Thus, in studies with a wide range of education it appears that the relationship with CDT is curvilinear.

The ages of participants included varied considerably across studies (minimum being 25 years in one study (Santana et al., 2013) to a maximum of 103 years in another study (Balduino et al., 2020).

None of the studies included quality of education as a measurement of education. Manly et al. (2002) suggested that years of education (which is used in all the included studies) is an inadequate measure of the educational experience among multicultural older populations and that including quality of education could improve the specificity of neuropsychological measures. QATSO analysis highlighted methodological flaws relating to this in all of the studies.

CONCLUSION

A consistent finding was that CDT performance is associated with level of education or literacy. The evidence suggests that the association with level of education is stronger at the lower levels of education and then plateaus as educational levels reach the minimum levels seen in many of the world's high-income countries.

Limitations

The studies included were carried out in a wide range of locations, but most of the studies were carried out in higher income countries. None of the included studies were carried out in LMICs where literacy remains lower, particularly amongst older adults.

Study quality was evaluated using an adapted version of the QATSO, which has not undergone a detailed examination of its reliability. However, on the sample of seven papers independently rated by two raters, scoring was shown to be consistent. (Appendix 1.1.3).

Implications for future research

This systematic review suggests that education influences performance of the CDT. However, very little research was carried out in lower income countries where the range of levels of education will vary the most. It would therefore be helpful to carry out further research in LMICs where there is an increasing life expectancy, with a significant number of adults, particularly older adults, having low levels of education and literacy.

It would be helpful to establish consensus standards for the most effective scoring system for the CDT. The review of Pinto and Peters (2009) provides sensitivity and specificity data for different scoring systems and this work could be developed as a basis for consensus.

One implication of there being a significant association between education and CDT performance is that normative data for the CDT will need to be adjusted to take account of educational level, with either stratification of normative data or use of regression-based norms. Another option is to explore whether other tasks could be developed that assess similar cognitive domains to the CDT (visuospatial, executive, praxis etc.), without the need for familiarity with clocks or numbers or the ability to use a pen or pencil so that they are less likely to be influenced by education.

References of studies included in SR

1. Ainslie, N.K., & Murden, R.A. (1993). Effects of education on the clock drawing dementia screen in non demented elderly persons. *Journal American Geriatric Society, 41*, 249-252.

2. Balduino, E., Alvarez Ribeiro de Melo, B., de Sousa Mota da Silva, L., Eduardo Martinelli, J., & Cecato, J.F (2020).The "SuperAgers" construct in clinical practice: neuropsychological assessment of illiterate and educated elderly. *International Psychogeriatrics, 32*(2), 191–198.

3. Bozikas, V.P, Giazkoulidou, A., Hatzigeorgiadou, M., Karavatos, A., Kosmidis, M.H. (2008). Do age and education contribute to performance on the clock drawing test? Normative data for the Greek population. *Journal of Clinical and Experimental Neuropsychology, 30,*199–203.

4. Caffarra, P., Gardini, S., Zonato, F., Concari, L., Dieci, F., Copelli, S., Freedman, M., Stracciari, A., & Venneri. (2011). Italian norms for the Freedman version of the Clock Drawing Test. *Journal of Clinical and Experimental Neuropsychology, 33*, 982–988.

5. De Noronha, Í.F.C.D., Barreto, S.D.S., & Ortiz, K.Z. (2018). The influence of education on performance of adults on the Clock Drawing Test. *Dementia & Neuropsychologia, 12*(1).

6. Fabricio, A.T., Aprahamian, I., Yassuda, M.S. (2014). Qualitative analysis of the Clock Drawing Test by educational level and cognitive profile. *Arq Neuropsiquiatr, 72*,289-295.

7. Hubbard, E.J., Santini, V., Blankevoort, C.G., Volkers, K.M., Barrup, M.S., Byerly, L., Chaisson, C., Jefferson, A.L., Kaplan, E., Green, R.C., Stern, R.A.(2008). Clock drawing performance in cognitively normal elderly. *Archives Clinical Neuropsychology, 23*, 295-327.

8. Kim, H., & Chey, J. (2010) Effects of education, literacy, and dementia on the clock drawing test performance. *Journal of International Neuropsychology Society, 16,* 1138-1146.

9. Leung, J.C.W., Lui, V.W.C., & Lam L.C.W., (2005). Screening for early Alzheimer's disease in elderly Chinese patients using the Chinese Clock Drawing Test. *Hong Kong J Psychiatry,15,* 14–17.

10. Marcopulos, B.A., McLain, C.A., Giuliano, A.J. (1997). Cognitive impairment or inadequate norms? A study of healthy, rural, older adults with limited education. *Clinical Neuropsychology, 11*,111-131.

11. Merims, D., Ben Natan, M., Milawi, D., & Boguslavsky, T. (2018). The clock-drawing test: Normative data in adult and elderly Israeli Arabs. *Psychogeriatrics, 18*(3), 175–181.

12. Nitrini, R., Caramelli, P., Herrera, E., Porto, C.S., Charchat-Fichman, H., Carthery, M.T., Takada, L.T., Lima, E.P. (2004) Performance of illiterate and literate nondemented elderly subjects in two tests of long-term memory. *Journal of International Neuropsychology Society, 10*, 634-638.

13. Paganini-Hill, A., Clark, L.J., Henderson, V.W., Birge, S.J. (2001) Clock Drawing: Analysis in a Retirement Community. *Journal American Geriatrics Society, 49:*941-947.

14. Santana, I., Duro, D., Frietas, S., Alves, L., & Simoes, R. (2013). The Clock Drawing Test: Portuguese Norms, by Age and Education, for Three Different Scoring Systems. *Archives of Clinical Neuropsychology, 28:*375-387.

15. Seigerschmidt, E., Mosch, E., Siemen, M., Forstl, H., & Bickel, H. (2002) The clock drawing test ad questionable dementia: reliability and validity. International Journal of Geriatric Psychiatry, 17, 1048–1054.

16. Shao, K., Dong, F-M., Guo S-Z., (2020). Clock-drawing test: Normative data of three quantitative scoring methods for Chinese-speaking adults in Shijiazhuang City and clinical utility in patients with acute ischemic stroke. Brain and Behaviour, 10.

17. Shanhu, X., Linhui, C., Xiaoqing, J., Jing, Y., Saizhu, X., Ying, X., Caixia, L., & Yu J. (2019) Effects of age and education on clock-drawing performance by elderly adults in China, The Clinical Neuropsychologist, 33(1), 96-105.

18. Siciliano, M., Santangelo, G., D'Iorio, A., Basile, G., Piscopo, F., Grossi, D., & Trojano, L. (2016) Rouleau version of the Clock Drawing Test: age- and education-adjusted normative data from a wide Italian sample. The Clinical Neuropsychologist, 30(1), 1501-1516.

19. Sugawara, N., Yasui-Furukori, N., Umeda, T., Sato, Y., Kaneda, A., Tsuchimine, S., Kaneko, S.(2014). Clock drawing performance in a community-dwelling population: Normative data for Japanese subjects. Aging & Mental Health, 14(5), 587–592.

20. Turcotte,V., Gagnon,M.-E., Joubert, S.,Rouleau, I., Gagnon, J.-F., Escudier, F.,Hudon, C. (2018). Normative data for the Clock Drawing Test for French-Quebec mid- and older aged healthy adults. The Clinical Neuropsychologist, 32(1), 91–101.

21. Von Gunten, A., Ostos-Wiechetek, M., Brull, J., Vaudaux-Pisquem, I., Cattin, S., Duc, R. (2008). Clock-drawing test performance in the normal elderly and its dependence on age and education. European Neurology, 60:73-78.

REFERENCES

Alzheimer's Disease International. (2010). World Alzheimer Report, London. Retrieved from https://www.alzint.org/u/WorldAlzheimerReport2010.pdf

Babins, L., Slater, M.E., Whitehead, V., & Chertkow, H. (2008). Can an 18-point clockdrawing scoring system predict dementia in elderly individuals with mild cognitive impairment? *Journal of Clinical and Experimental Neuropsychology, 30,* 173-86.

Borson, S., Brush, M., Gil, E., Scanlan, J., Vitalino, P., Chen, J., Cashman, J., Sta Maria, M.M., Barnhart. R., & Roques. J. (1999). The Clock Drawing Test: utility for dementia detection in multiethnic elders. *Journal Gerontological Medical Science,54,* 534– 540.

Brodaty, H. & Moore, C.M. (1997). The Clock Drawing Test for dementia of the Alzheimer's type: a comparison of three scoring methods in a memory disorders clinic. *International Journal of Geriatric Psychiatry,12,* 619–627.

Cahn D.A., Salmon, D.P., Monsch, A.U., Butters, N., Wiederholt, W.C., & Corey-Bloom, J. (1996) Screening for dementia of the Alzheimer type in the community the utility of the Clock Drawing Test. *Archives Clinical Neuropsychology,11,*529–539.

Freedman, M., Leach, L., Kaplan, E., Winocur, G., Shulman, K.I., Delis, C.D. (1994) *Clock Drawing: a neuropsychological analysis.* New York: Oxford University Press.

Freund, B., Gravenstein, S., Ferri, R., Burke, B.L. (2005) Drawing clocks and driving cars. *Journal of General Internal Medicine, 20,* 240-4.

Hsieh, S., Schubert, S., Hoon, C., Mioshi, E., & Hodges, J.R. (2013). Validation of the Addenbrooke's Cognitive Examination III in Frontotemporal Dementia and Alzheimer's Disease. *Dementia Geriatric Cognitive Disordors,36,* 242-250.

Kim, H., & Chey, J. (2010) Effects of education, literacy, and dementia on the clock drawing test performance. *Journal International Neuropsychological Society, 16,* 1138-1146.

Kim, S., Jahng, S., Yu, K. H., Lee, B. C., & Kang, Y. (2018). Usefulness of the clock drawing test as a cognitive screening instrument for mild cognitive impairment and mild dementia: An evaluation using three scoring systems. *Dementia and Neurocognitive Disorders, 17*(3), 100–109.

Lam L.C., Chiu, H.F., Ng, K.O., Chan, C., Chan, W.F., Li, S.W., Wong, M. (1998). Clock-face drawing, reading and setting tests in the screening of dementia in Chinese elderly adults. *Journal of Gerontology, 53(*6)353-7.

Libon, D.J., Malamut, B.L., Swenson, R., Sands, L.P., & Cloud, B.S. (1996) Further analyses of clock drawings among demented and nondemented older subjects. *Archives of Clinical Neuropsychology, 11,* 193–205.

Manoj, M., Crawford, S., Dutt, A., Ghosh, A., Nandi, R., and Evans, J.J. (2015) Pizzas and Papadums: An alternative to the Clock Drawing Test for the assessment of cognition. Poster presented at the International Neuropsychological Society 2015 Mid-Year Meeting / 5th ASSBI Pacific Rim Conference July 1-4, 2015 Sydney, Australia.

Manos, P.J., Wu, R. (1994). The ten point clock test: a quick screen and grading method for cognitive impairment in medical and surgical patients. *International Journal Psychiatry, 24,* 229–244.

Mendez, M.F., Ala, T., Underwood, K.L. (1992). Development of scoring criteria for the clock drawing task in Alzheimer's disease. *Journal American Geriatric Society, 40*(11),1095-9.

Montani, C., Bouati, N., Pelissier, C., Couturier, P., Jasso-Mosqueda, G., Hugonot, R., Franco, A. (1997). Cotation et validation du test du cadran de l'horloge en psychométrie chez le sujet âgé. *Encéphale,23,*194–199.

Nishiwaki, Y., Breeze, E., Smeeth, L., Bulpitt, C. J., Peters, R., Fletcher, A. E. (2004). Validity of the Clock-Drawing Test as a screening tool for cognitive impairment in the elderly. *American Journal of Epidemiology, 160,* 797-807.

Peterson, A. (2005). On the Use of Beta Coefficients in Meta-Analysis. *Journal of Applied Psychology, 90*(1) 175–181.

Pinto, E., & Peters, R. (2009). Literature review of the Clock Drawing Test as a tool for cognitive screening. *Dementia Geriatric and Cognitive Disordors,27*(3) 201-13.

Rouleau, I., Salmon, D.P., Butters, N., Kennedy, K.C., McGuire, K.(1992) Quantitative and qualitative analyses of clock drawings in Alzheimer' and Huntington' disease. *Brain and Cognition,18,* 70-87.

Royall, D.R., Cordesa, J.A., & Polka, M.(1998) Clox: an executive clock drawing task. *Journal Neurological Neurosurgical Psychiatry, 64,* 588–594.

Shulman, K. I. (2000). Clock-drawing: Is it the ideal cognitive screening test? *International Journal of Geriatric Psychiatry, 15,* 548-561.

Shulman, K.I., Gold. D.P., Cohen, C.A., & Zucchero, C.A. (1993). Clock-drawing and dementia in the community: a longitudinal study. *International Journal Geriatric Psychiatry, 8,* 487-96.

Sunderland, T., Hill, J.L., Mellow, A.M., Lawlor, B.A., Gundersheimer, J., & Newhouse, P.A. (1989). Clock drawing in Alzheimer's disease: a novel measure of dementia severity. *Journal American Geriatric Society, 37,* 725-9.

Tripathi, R. K., Verma, Y., Srivastava, A., Shukla, T. S., Usman, K., Ali, W., & Tiwari, S. C. (2020). Usefulness of clock-drawing test in Indian older adults with diabetes mellitus. *Indian journal of psychiatry, 62*(1), 59–65.

Todd, M.E., Dammers, P.M., Adams, S.G., Todd, H.M., & Morrison M (1995). An examination of a proposed scoring procedure for the clock drawing test reliability and predictive validity of the clock scoring system (CSS). *American Journal of Alzheimers Diseases, 4,* 22–26.

UNESCO Institute of Statistics (2017, September 1). Literacy Rates Continue to Rise from One Generation to the next. Retrieved from http://uis.unesco.org/en/topic/literacy

United Nations, Department of Economic and Social Affairs, Population Division (2017). World Population Ageing. United Nations Copyright. Retrieved from: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Report.pdf

Watson, Y.I, Arfken, C.L, & Birge S.J.(1993). Clock completion: an objective screening test for dementia. *Journal American Geriatric Society, 41,*1235-40.

Wolf-Klein, G.P., Silverstone, F.A., Levy, A.P., & Brod, M.S. (1989). Screening for Alzheimer's disease by clock drawing. *Journal American Geriatric Society, 37,* 730-4.

Wong, W.C., Cheung, C.S., Hart, G.J. (2008). Development of a quality assessment tool for systematic reviews of observational studies (QATSO) of HIV prevalence in men having sex with men and associated risk behaviours. *Emergency Themes Epidemiology, 5*, 23.

World Health Organization and Alzheimer's Disease International. (2012). Dementia a public health priority. Retrieved from https://www.who.int/mental_health/publications/dementia_report_2012/en/

CHAPTER 2: EXAMINATION OF THE VALIDITY OF THE 'PAPADUM TEST': AN ALTERNATIVE TO THE CLOCK DRAWING TEST FOR PEOPLE WITH LOW LEVELS OF EDUCATION/LITERACY

Crombie, M.*, Dutt, A., Dey, P., Nandi, R & Evans, J.J. (2021)

***Address for Correspondence:**

Institute of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

1st Floor, Administration Building,

Gartnavel Royal Hospital

1055 Great Western Road

Glasgow, G12 0XH

Tel: 0141 211 0607

Email:

Word Count (including tables, figures, references. Excluding Plain English Summary): 6657

Prepared in accordance with guideline for submission to "The Clinical Neuropsychologist"- (See appendix 1)

PLAIN ENGLISH SUMMARY

Background

The prevalence of dementia is increasing in low/middle income countries (LMICs) due to increases in life expectancy. Reliable ways to assess for difficulties that may indicate dementia are required. Well established tools include the Clock Drawing Test (CDT). However, a certain level of education and familiarity with writing numbers are required to complete this task. As LMICs have high rates of illiteracy and low education, it is essential that we have tasks that adequately assess for dementia in these populations. The Pizza/Papadum test was developed as an alternative, as it is believed that this could provide similar information but without the requirement to read or write.

Aims

Extending previous research (Manoj et al., 2015), this study looked at the usefulness of the Papadum Test in an Indian population. We assessed the impact years of education had on performance on the CDT and the Papadum test and how similar these tests were to each other. We also analysed how effectively each test distinguished people with dementia from people without dementia.

Method

89 healthy adults and 59 adults with a diagnosis of dementia from hospitals in Kolkata undertook the ACE III which included the CDT; participants were asked to draw a clock with all the numbers and place the hands at 10 past 5. Participants were then asked to complete the Papadum test; participants were given a circle of paper and asked to imagine it is a papadum and to show how it would be divided equally among six people. They were also asked to do this with an actual papadum.

Main findings and conclusions

We found that level of education influenced CDT performance but did not affect Papadum test performance. This suggests that CDT is not suitable for those with low levels of literacy or education. In healthy controls the CDT and Papadum tests were not closely associated. However, the Papadum test and CDT were similar in their ability to distinguish controls and people with dementia. Therefore, we conclude that the Papadum test could provide an alternative as a screening tool to the CDT for use with people who are illiterate or have low levels of education. Further studies are required, however.

ABSTRACT

Objectives The clock drawing test (CDT) is a widely used cognitive screening test. However, CDT performance is affected by education. This study examined an alternative: the Papadum test, designed for people with low levels of education/literacy. The association between education and test performance, correlation between CDT and Papadum test, and diagnostic accuracy of both CDT and Papadum tests were examined.

Method 89 healthy literate adults and 59 literate adults with a diagnosis of dementia from hospitals in Kolkata, India undertook the CDT and the Papadum test.

Results Education had a significant association with the CDT but not with the Papadum test. Across the whole sample there was a significant correlation between CDT and Papadum, but not within separate groups of healthy controls and patients. Diagnostic accuracy for the Papadum Test was similar to that for CDT.

Conclusions Results highlight the strong influence that education has on CDT performance indicating that it is not suitable for those with low levels of literacy. The Papadum test could provide a viable alternative as a screening tool to the CDT for use with people who are illiterate or have low levels of education. Further validation studies are required.

INTRODUCTION

An increase in life expectancy is occurring across the world, with the greatest increase being seen in low-middle income countries (LMICs) (WHO, 2012). It is expected that the number of over 80s will grow significantly by 2050 in several countries. Additionally, there is expected to be an increase in the number of people living with age related diseases, including dementia (Ferri et al., 2005; Prince et al., 2013). It is estimated that by 2040, 71% of 81.1 million dementia cases will be in lower income countries (Ferri, 2005).

Dementia is considered as one of the leading causes of disability among older people and typically contributes to many systemic and socioeconomic difficulties (WHO, 2012). In 2015, the total global societal cost of dementia was estimated to be $818 billion. The World Health Organization (WHO, 2012) recognises dementia as a public health priority. There is a clear need for effective neuropsychological testing that can be used to help screen for dementias to allow for timely diagnosis and intervention within low-middle income populations.

The Clock Drawing Test (CDT) was devised in the 20th century and has been widely used in research protocols and in clinic for screening individuals suspected of having cognitive impairment or dementia (Royall et al., 1998; Critchley, 1953). In the last thirty years, an evidence base has formed for its use as an early screening tool of cognitive impairment, especially in Alzheimer's disease (Scanlan et al., 2002). The CDT makes demands on a range of cognitive domains; memory, attention, visual memory, reconstruction, planning, motor skills, visuospatial abilities, concentration, and auditory comprehension (Royall et al., 1998). It is a brief tool that has been widely applied internationally and involves asking the client to draw or copy the face of a clock and then to draw the hands to indicate a particular time. The CDT is included as a subtest in the Addenbrooke's Cognitive Examination (ACE III; Hsieh et al., 2013), which is a widely used cognitive screening tool. The ACE III is a relatively quick and easy to administer measure that assesses five cognitive domains; memory, attention, language, visuospatial and perceptual abilities.

The CDT is a very popular screening tool; however, it was developed to be used in Western, educated populations. One of the challenges in screening for dementia in LMICs is related to performance variances due to educational and cultural influences. It is widely documented that neuropsychological test performance may be considerably influenced by aspects such as culture, language, education, and literacy (Ardila et al., 1989; Ardila, 2005). Strong evidence has further developed over the last decade indicating that CDT results can be influenced by education (Kim and Chey, 2010; Nielsen & Jorgensen, 2013) which can lower the specificity of the test. Studies have discussed significant limitations in the use of the CDT with people with low levels of education or literacy. Liberman et al. (1999) found that poorer performance on the CDT has also been observed in those who do not speak English.

De Noronha et al. (2018) found that CDT performance was significantly impacted for illiterate individuals. Illiteracy refers to the inability to read or write a simple message (UNESCO, 2017). Literacy rates in LMIC vary among older generations and amongst individuals in rural areas who often receive limited or no formal education. Nielsen & Jorgensen (2013) further found that healthy illiterate individuals may experience problems with graphomotor construction when asked to engage in the CDT task. Another commonly used screening tool for cognitive impairment, the Mini-Mental State Examination (MMSE) has also been shown to be influenced by schooling, displaying a lack of validity in illiterate populations (Kalafat et al., 2003). As well as potentially having an influence on performance, low education and illiteracy are considered major risk factors for developing Alzheimer's disease (Stern et al., 1994) in LMICs.

Levels of literacy in India are particularly low; although the country has made significant progress in improving literacy over the years, it continues to be home to 313 million illiterate people (UNESCO, 2017). Tripathi et al. (2020) recently examined the usefulness of the CDT in screening Indian older adults for cognitive impairment and found that education and language are significant variables that correlate with CDT performance. Apart from difficulties with screening tools, other difficulties can exist regarding a general lack of awareness of dementia in the general population in India, in particular attributing common symptoms to 'normal aging' (Khan, 2011).

The present study examined the validity of a task designed to examine similar cognitive skills to the CDT, but which does not require the ability to read or write and therefore may serve as an alternative to the CDT for illiterate or low educated populations. This test can be adapted culturally and was designed to be an educationally unbiased alternative to the CDT. A study carried out in NHS Greater Glasgow and Clyde, Scotland previously used this test where it was referred to as 'the Pizza test' (Manoj et al., 2015). The task involved the person being provided with a paper circle and being asked to imagine it was a pizza. They were told that they had to divide the pizza between six people equally, so they had to fold and divide the paper into six equal pieces. The test is considered to assess cognitive abilities such as attention, planning, problem solving, visuo-spatial and praxis skills. Results suggested that this task captures similar cognitive domains to the CDT. No effects of education were observed suggesting that the task may be appropriate for those with low, or no education. Results also showed that the Pizza test can distinguish between patients with dementia and those without dementia with an overall good diagnostic accuracy.

The data for this present study were collected in Kolkata, India, where the 'Pizza' test was referred to as the 'Papadum' test. The study included the use of a paper 'papadum' and an actual papadum. Literate participants were included as they were able to attempt both the CDT and papadum to examine correlations between performance on the two tests. The impact of education on the CDT and the

Papadum test performance was examined. In addition, the diagnostic accuracy of the CDT and Papadum test was examined.

Hypothesis

It was hypothesised that education would have a significant association with the CDT but not with the Papadum test. Performance on the Papadum test (both paper and actual papadum versions) was predicted to be significantly correlated with the performance on the CDT.

Exploratory analyses were conducted to assess diagnostic accuracy of the Papadum test.

METHOD

Participants

A total of 59 literate patients clinically diagnosed with Alzheimer's disease (AD) or Vascular dementia (VaD) participated in the study. Patients with AD and VaD in the mild and moderate stages of the illness as evident from their scores on the Mini Mental State Examination (MMSE) (Folstein et al., 1975)/ Bengali Mental State Examination (BMSE) (Das et al., 2006) as well as their Clinical Dementia Rating (Hughes et al., 1982) scores were invited to participate in the study. The patients were seen at Duttanagar Mental Health Centre and at Apollo Gleneagles Hospitals, Kolkata, India. Patients were assessed by a neurologist or a psychiatrist and underwent an MRI or CT scan. Each patient underwent a comprehensive neuropsychological examination by an experienced neuropsychologist which included a range of neuropsychological tests. Clinicians making the diagnosis of dementia were blind to Papadum test scores. Patients with AD with a history of cerebrovascular disease or significant changes in the brain suggestive of cerebrovascular pathology or patients with severe Parkinson's Disease were excluded from the study. None of the patients included in the study had other neurological illnesses, history of psychiatric illness, head injury, major medical illness, or substance abuse.

A total of 89 cognitively healthy literate adults were also included in this study who all resided in Kolkata and were in the age range of 40 years and above. These participants were a) relatives or friends of patients attending the Neuropsychology and Clinical Psychology unit at Duttanagar Mental Health Centre, b) family members of other patients attending the hospitals, c) volunteer hospital staff, or d) members in the community (acquaintances of other participants who volunteered for the study). Inclusion criteria included having a minimum of one year of education, with a MMSE/BMSE score of above 25. The Hospital Anxiety and Depression Scale (HADS) was also administered. Exclusion criteria were applied; individuals with cognitive complaints, hearing or vision problems or any history of neurological or psychiatric illnesses were excluded from this study.

The primary language of all participants was Bengali. Demographic details of age, sex and years of education were gathered by a structured interview for each participant.

Ethical approval

Ethical approval for the study was obtained from the Institutional Ethics Committees of Apollo Gleneagles Hospitals, Kolkata and Duttanagar Mental Health Centre, Kolkata (see Appendix 2.1 & 2.2). Informed consent was sought and provided by all participants.

Materials

Addenbrooke's Cognitive Examination (ACE) III – Bengali Version (including CDT)

The ACE III – Bengali Version was administered to all participants. The ACE III is a brief cognitive screening tool that assesses five cognitive domains: attention, memory, verbal fluency, language, and visuospatial abilities. The maximum total score is 100. The ACE III includes the Clock Drawing Test as a subtest to assess visuospatial abilities. For the purposes of this paper, only scores of the CDT component will be reported.

Participants were asked to draw a clock including all the numbers and set the hands to ten past five.

The clocks were scored quantitatively according to two scoring systems based on the clockface, numbers and hands using the ACE III 5-point scoring system and the Rouleau 10-point scoring system (Rouleau et al., 1992) (See Appendix 2.3). This paper will focus on the analysis of the Rouleau system scores only.

Some of the scoring criteria were modified while using the Rouleau quantitative scoring system to capture culturally different information, for example, clock numbers written partly in English script and partly in Bengali. Each clock was scored independently by two members of the research team.

Papadum Test

The Papadum test was administered to all the participants. Two versions of this test were used;

a) Paper papadum: The individual was given a circular piece of paper measuring 18cm in diameter. The instructions were: 'Imagine that this is a papadum which you have to share amongst six members in your family. Could you kindly show me how you will tear the paper so that the six members in your family get an equal share'.

b) Actual papadum: A dried / unfried papadum was given to each participant. They were asked to divide it into six equal slices. The instructions were: 'Imagine that you have to share this papadum amongst six members in your family. Could you kindly show me how you will tear the papadum so that the six members in your family get an equal share'.

See Appendix 2.4 for scoring criteria. The maximum score was 18. Each papadum was scored independently by two members of the research team.

Statistical analysis

Scoring reliability for the CDT and Papadum tests was examined. Two members of the research team rated all the CDT and Papadum performances. Where discrepancies occurred between raters, a third senior member of the research team was consulted, and a final score was determined. An example of where there were discrepancies between raters was in relation to interpretation of the threshold for what constituted a 'rotation' of numbers on the clock face.

We examined whether there was a correlation between scores on the Papadum Test with the CDT scores within the whole sample, healthy controls and patients. As the data were not normally distributed, Spearman's Correlations were carried out. The effects of education, age and sex on performance on both the CDT and Papadum tests were examined to determine whether demographic factors are associated with test performance. Regression analyses were also carried out. A Kruskal Wallis test was also conducted to compare levels of education with performance on the tests.

The diagnostic accuracies of the tests were examined using Receiver Operating Characteristic (ROC) curves for the data from the patients and healthy controls.

RESULTS

Demographic Characteristics

One hundred and forty eight adults (73 males and 75 females) were included in the study and analysed. Of this total, 89 were healthy controls and 59 were patients with a diagnosis of dementia (30 had AD, whilst 29 had VaD). Mean age (years) and education (years) of the total sample were 63.24 (SD = 11.46; max 87, min 40) and 11.22 (SD = 5.01, max 21, min 1) respectively. With a total of 148 participants, the correlation analysis had 80% power to detect correlation as small as r=0.225.

The average age in years of the patients (23 females, 36 males) was 69.69 (SD 8.65; max 87, min 48) with average education in years as 13.68 (SD 3.62; max 21, min 4).

The average age of healthy controls (50 females, 39 males) was 58.97 (SD 11.12; max 86, min 40). Average education level in years was 9.58 (SD 5.16; max 21, min 1).

The healthy controls were significantly younger than the patients ($t(142.2) = -6.579$, $p < 0.001$, $d= 1.05$). The healthy controls also had a significantly lower level of education than the patients ($t(145.5)=-5.674$ $p<0.001$, $d=.89$). There was a significant difference in terms of numbers of males/females between the groups (χ^2 (1) = 4.20, $p<0.04$).

The data for each test for all participants (inclusive of patient and healthy controls), as well as patients and healthy controls separately are included in Tables 1-3 respectively.

Table 1: Actual Papadum, Paper Papadum and CDT test data for all participants

Tests	Mean	SD	Minimum	Maximum	Max.possible score
Actual Papadum	12.47	4.10	0	18	18
Paper Papadum	12.83	5.08	0	18	18
CDT	6.76	2.68	0	10	10

Table 2: Actual Papadum, Paper Papadum and CDT test data for patients

Tests	Mean	SD	Minimum	Maximum	Max.possible score
Actual Papadum	10.03	3.85	0	18	18
Paper Papadum	9.83	4.97	0	18	18
CDT	5.08	2.52	0	10	10

Table 3: Actual Papadum, Paper Papadum and CDT test data for healthy controls

Tests	Mean	SD	Minimum	Maximum	Max.possible score
Actual Papadum	14.09	3.41	6	18	18
Paper Papadum	14.81	4.12	0	18	18
CDT	7.87	2.17	1	10	10

Inter-rater Reliability

Regarding inter-rater reliability of papadum and CDT scoring, there was 76% agreement between raters for the papadum scores, 86% agreement for the paper scores and 80% agreement for the CDT scores. Scoring challenges emerged for the CDT in applying the Rouleau method and therefore adaptions/clarifications were agreed; for an error in "rotation of numbers", six numbers or more rotated numbers were required to be interpreted as rotated. For the scoring of the Papadum tests, discrepancies emerged in relation to what constituted as a triangular shape; it was agreed that a minimum ratio of 2:1 regarding the top of the section of Papadum being at least twice the length of the bottom part was required to consider the Papadum to be "triangular" in shape.

To examine the distributions of the data, histograms and box plots were used to visually analyse whether the data were distributed normally. The Kolmogorov-Smirnov Test to assess the normality in the study data (actual papadum, paper papadum and CDT) was also used. The following results were obtained; $D(148)=.126$, $p<0.001$ for the actual papadum test, $D(148)=.203$, $p<0.001$ for the paper papadum test and $D(148)=.206$, $p<.001$ for the CDT. Therefore, data did not follow a normal distribution and nonparametric analyses were used.

Education and Performance on the Tests in Healthy Controls

The relationship between education and performance on CDT, Papadum and Paper within the healthy controls was examined using Spearman Correlation. The correlation co-efficient for actual papadum was $rs=-.010$, $p=.927$; for the paper papadum it was $rs=.054$, $p=0.618$; and for CDT it was $rs=0.507$, $p<.001$. Thus, there was a significant relationship between education and CDT performance but there was no association between either paper or actual papadum and education.

A Kruskal Wallis test was also conducted to compare the performance of the healthy controls across three levels of education (Level 1: 1 – 4 years, Level 2: 5-8 years, Level 3: 9-12 years, Level 4: 13+ years) on the three tests. The results displayed that there was no significant difference between the four levels of education on the actual papadum test $H(3) =1.479$, $p=.687$, or the paper papadum test $H(3) =1.243$, $p=.743$. There was a significant relationship between the CDT and the four education levels $H(3)$ 27.287, $p<.001$.

There was a significant relationship between Level 1(1-4 years) and the three more highly educated levels when specific comparisons were made for the CDT; ($U=108$, $p<.020$), ($U=69.5$, $p<.001$) and ($U=75$, $p<.001$) respectively. When comparing Level 2 (5-8 years) with Level 3 (9-12 years) ($U=130.5$, $p<.05$) and with Level 4 (13+ years) ($U=136.5$, $p=.002$), a significant difference was also found. There was no significant difference between Level 3(9-12 years) and 4 (13+ years) however ($U=243$, $p=.205$).

Figure 1 presents a scatterplot of education (in years) against CDT score to illustrate the nature of the relationship in the healthy controls. The plot shows that there is a curvilinear relationship such that CDT scores increase with education up to about 8 years, and then plateaus.

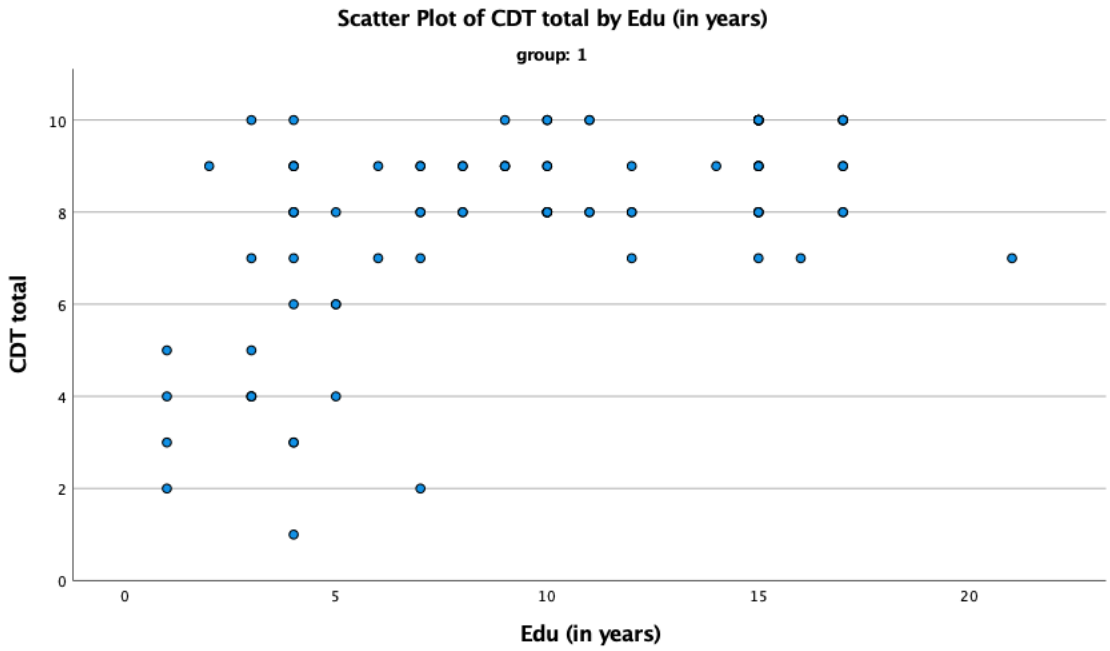


Figure 1: Scatterplot of education (in years) against CDT total score in healthy controls

This was explored in a regression analysis that included CDT as the dependent variable and education in an initial model and then education squared was added to the model. The results confirmed that addition of education squared increased the predictive power of the model (Model 1, F (1,87) = 39.25, p<.001, R^2=.311; Model 2, F(2, 86) = 29.16, p<0.001, R^2=.390, R^2change=.093, Fchange (1,86) =13.45, p<0.001).

Age & Performance on Tests in Healthy Controls

The relationship between age and performance on CDT, actual papadum and paper papadum within the healthy controls was examined using Spearman correlation. The correlation co-efficient for actual papadum was rs= .014, p=.898, for paper papadum was rs= -.017, p=.876 and for CDT rs=-.054, p=.612. Therefore, there was no significant relationship with age.

Sex & Performance on Tests in Healthy Controls

There was no significant effect of sex for the actual papadum (U=932, Z=-.357, p=0.721) or paper papadum (U=875, Z=-.898, p=.369) amongst the healthy controls. For the CDT, females had lower scores though the difference was not significant, with a small-medium effect size (U=744, Z=-1.954, p=0.051, r=0.21). The female healthy controls had a lower level of education and in a regression model

with CDT as the dependent variable and education and sex as predictors, only education was significantly associated with CDT performance (beta=.539, t=5.941, p<0.001), whilst sex was not a significant predictor (beta=-.097, t=-1.069, p=.288).

Relationship between Tests for total sample, healthy controls and patient group

The Spearman correlation for the total sample (healthy controls and patients) between the actual papadum and CDT was rs=.221, p=.007; the paper papadum and CDT was rs=.304, p<.001 and the correlation between actual papadum and paper papadum was rs=.582, p=<.001.

However, there was no significant relationship between CDT and papadum tests when looking at the healthy and patient groups separately. The Spearman correlation for healthy controls between the actual papadum and the CDT was rs=-.052, p=.626; the paper papadum and the CDT was rs= -.038, p=.724 and the correlation between the actual papadum and the paper papadum was rs =.486, p <.001. A regression analysis with CDT as dependent variable and paper papadum, education, age and sex as predictors found that education was a significant predictor (beta=0.565, t=6.083, p<0.001) but paper papadum was not (beta=-.056, t=-.625, p=.534) and nor was sex (beta=-.116, t=1.248, p=.215) or age (beta=-.112, t=-1.204, p=.232). A similar result was obtained when using the actual papadum score, education, age and sex as predictors, such that education was a significant predictor (beta=.559, t=6.051, p<0.001) but actual papadum was not (beta=-.082, t=-.926, p=.357) and nor was sex (beta=-.122, t=-1.317, p=.191) or age (beta=-.115, t=-1.229, .223).

For the patient sample; The Spearman correlation for actual papadum and CDT was .041, p=.760; paper papadum and CDT was .184, p=.164 and the correlation for actual papadum and paper papadum was .379, p=.003. A regression analysis with CDT as dependent variable and paper papadum, age, education and sex as predictors found that sex was a significant predictor (beta=-.344, t=2.517, p=0.015) but paper papadum (beta=.102, t=.836, p=.407), age (beta 0.39, t=.319, p=751) and education (beta=.152, t=1.120, p=.268) were not. A similar result was obtained when using the actual papadum score, age, sex and education as predictors, such that sex was a significant predictor (beta=-.346, t=2.517, p=0.015) but actual papadum (beta=.037, t=.301, p=.765), age (beta=0.52, t=4.16, p=.679) and education were not (beta=.160, t=1.175, p=.245). This was further explored by examining dementia severity data from clinician ratings on the Clinical Dementia Rating Scale (CDR; Hughes et al., 1982). For the patient group the median score was 1, interquartile range 1. Females (Median =2) were found to be more severely impaired than males (Median =1) (U=276, Z=-2.335, p=.020) suggesting that gender was associated with CDT performance as a result of differences in dementia severity.

As the patients and control groups differed in mean age, length of education and sex, the groups were compared with a binary logistic regression, with group as dependent variable, and age, education and sex as covariates along with actual Papadum, paper Papadum or CDT scores in three separate models. Results showed that each of the tests significantly distinguished the groups over and above differences

in age/education/sex. The logistic regression model for CDT was statistically significant (χ^2 (3) =111.47, p<0.001, Nagelkerke R^2=.716), with CDT score being a significant predictor of group membership (p<0.001), correctly classifying 87.8% of participants. Similarly, the model for the actual Papadum was significant (χ^2 (3) =72.07, p<0.001, Nagelkerke R^2=.521), with actual Papadum score being a significant predictor of group membership (p<0.001), correctly classifying 83.8% of participants, and finally the model for the paper Papadum was also significant (χ^2 (3) =76.76, p<0.001, Nagelkerke R^2=.547) with paper Papadum score being a significant predictor of group membership (p<0.001), correctly classifying 80.4% of participants.

Diagnostic Accuracy

Receiver Operating Characteristic (ROC) curves for the actual papadum (Figure 2), paper papadum (Figure 3) and CDT scores (Figure 4) differentiating participants with dementia and healthy controls were constructed. The area under the curve (AUC) for the CDT was .794, the optimal cut off score for the CDT in this study was found to be 6.5 which displayed a sensitivity of .712 and specificity of .82 with Youden's index as 0.532 and LR + 2.847 and LR- 0.253. Positive and negative predictive values were PPV (0.811) and NPV (0.724). For scoring of the CDT, Rouleau et al. (1992) stated ≤7 as cut off and Duro et al. (2018) found ≤7also.

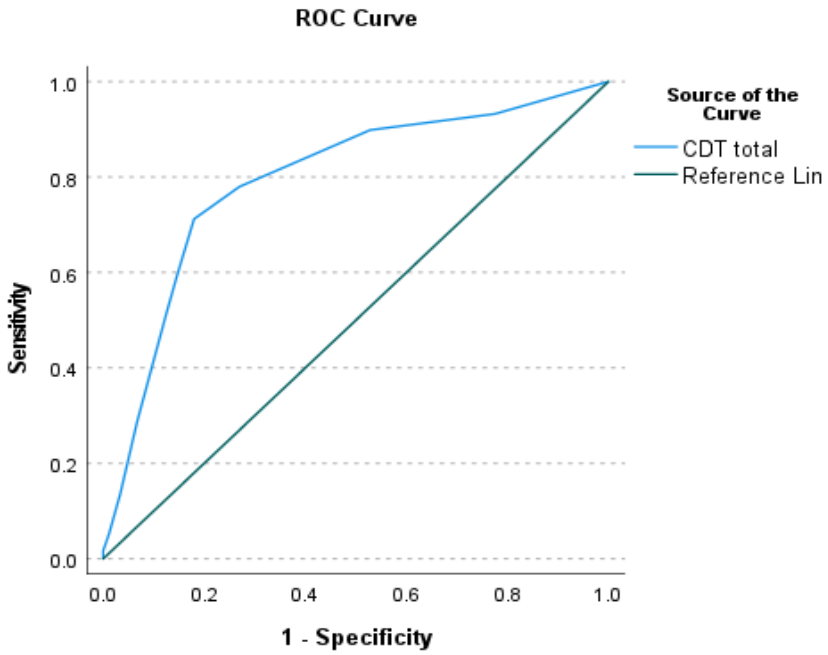


Figure 2 ROC curve for CDT

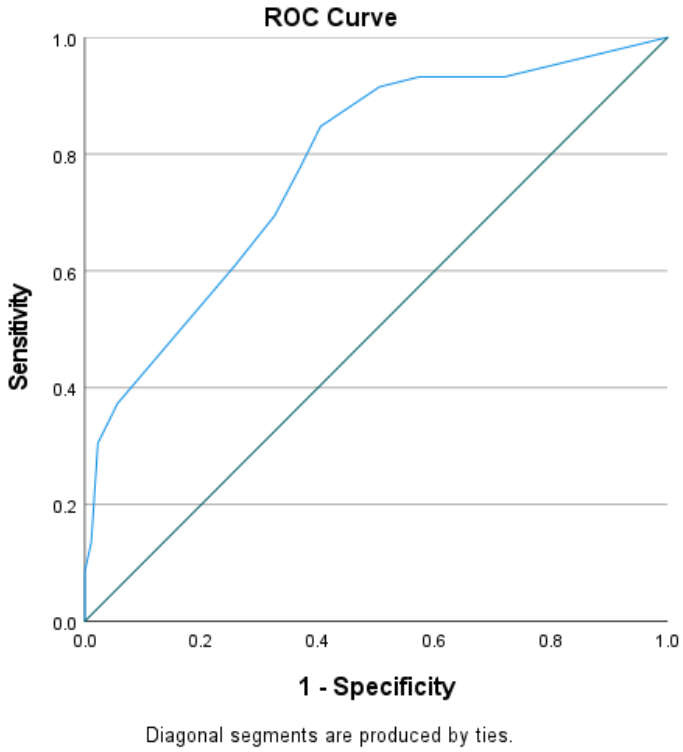


Figure 3 ROC curve for Actual Papadum

The AUC foı the actual Papadum was .778, the optimal cut off score for the Papadum in this study was found to be 13.5 with a sensitivity of .847 and specificity of .596 with Youden's index as 0.443 and LR + 3.895 and LR- 0.477. Positive and negative predictive values were PPV (0.855) and NPV (0.581).

An alternative approach to selecting cut-offs is to focus on a definition of impairment based on percentiles. For the actual papadum in the healthy controls, only 2.2% of the sample scored 8 or less and this therefore may be a useful means of defining impaired performance. With this cut off, 30.5% of the patient sample scored 8 or less.

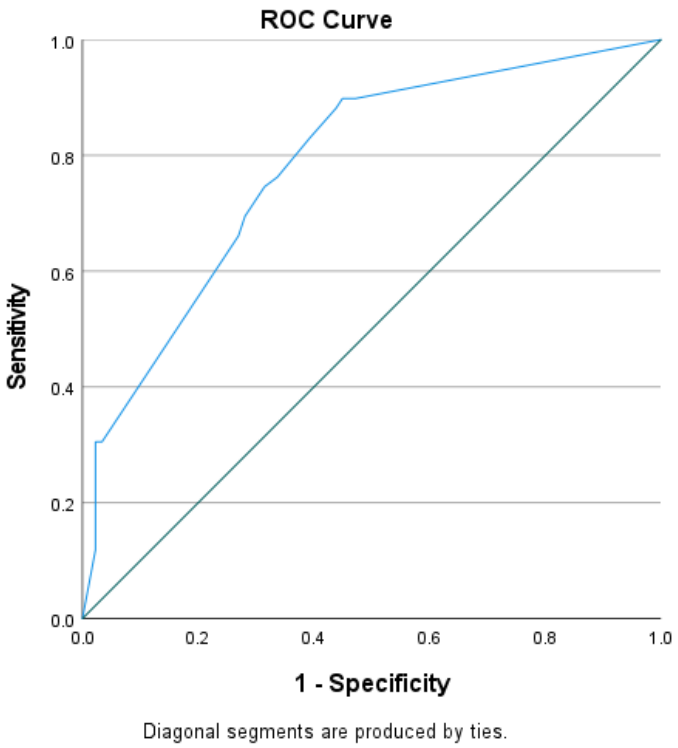


Figure 4- ROC curve for paper papadum

The AUC for the Paper papadum was .782, the optimal cut off score for the Paper papadum version in this study was found to be 15.5 with a sensitivity of .881 and specificity of .562 with Youden's index as 0.443 and LR + 4.723 and LR- 0.497. Positive and negative predictive values were PPV (0.877) and NPV (0.571). Manoj et al. (2015)'s Pizza study reported a cut-off of 13/18 and was previously found to show maximum Youden index of 0.50 with a sensitivity of 0.643 with a specificity of 0.857, AUC 0.802.

Alternatively, selecting a cut-off based on percentiles; for the paper papadum in the healthy controls, only 3.4% of the sample scored 9 or less and this therefore may be a useful means of defining impaired performance. With this cut off, 31% of the patient sample scored 9 or less.

DISCUSSION

This study shows that CDT performance is impacted by level of education. The association with education was strong in the lower end of educational range (up to about 8 years of education) and then reduces as educational levels increase, suggesting that the test may be useful for those with higher levels of education but is problematic in relation to use with people with less than about eight years of education. This finding is consistent with several other studies in other parts of the world that have demonstrated the effect of education on CDT performance (Kim & Chey, 2010; Ainslie & Murden,

1993; Crombie & Evans, submitted) meaning that caution must be used if this test is used as a screening tool for cognitive impairment or dementia in less educated or illiterate populations.

The aim of the development of the Papadum/Pizza test was to capture cognitive abilities such as attention, planning, problem solving, praxis skills, visuo spatial and visuo constructive abilities, similar to those assumed to be required for effective performance of the CDT. The aim of the Papadum/Pizza test was, however, to not have the requirement for a certain level of literacy or experience in the use of writing implements. The Papadum/Pizza test is quickly administered, taking on average two to three minutes. It does not involve extensive training to administer or score.

The study highlights the challenge of taking a test scoring system designed in a Western, English-speaking context and using it in a different cultural and linguistic context. The Rouleau CDT scoring system was adapted to address the common inclusion by healthy controls of numbers on the clock in different scripts (i.e. Bengali and English). It also highlighted the importance of having a robust reliability analysis process to ensure consistency of scoring between raters. Despite using a well-established scoring system there were examples of where well-trained raters gave different scores, requiring adjudication from additional senior researchers. For this study this process ensured consistency of scoring between the dementia and healthy control groups. But it raises questions regarding the reliability of scoring of tasks such as the CDT in clinical practice, particularly in a cross-cultural context where issues may arise that would not have occurred in the well-educated, mono-lingual context in which the test scoring system was originally developed.

No association was found between education and the paper or actual Papadum tests. This was what was expected and supports the idea that the Papadum/Pizza test may be useful for those with low levels of education.

There were significant correlations between both versions of the Papadum (actual and paper) test and the CDT when analysing the sample of healthy controls and patients combined. However, within just the healthy controls (and within the patient group) correlations were very low. A stronger relationship between CDT and the Pizza test was reported in Manoj et al.'s (2015) study. In the present study, there was a much greater range of education which may have been impacting on the CDT performance to a much greater extent than in Manoj et al.'s study. Another possibility is that the cognitive skills required for the CDT are different to those needed to complete the Papadum test. However, it is more likely that there is overlap in the cognitive demands of each test, but the relative contribution of various cognitive skills varies between the tasks, reducing the precision of any association between test performance. Another relevant factor is that screening tools such as the CDT typically have a highly skewed distribution of scores as they are relatively 'easy' tasks that are completed perfectly by a large proportion of participants. In our healthy control sample, although there was a wide range of scores, 73% of participants scored within the 8-10 point range (max score is 10). Similarly, for the papadum

tests there was also a full range of scores obtained but for the actual papadum test, 28.1% scored the maximum score of 18, and for the paper papadum 52.8% scored the maximum score.

We did not find a significant relationship between age and CDT or Papadum Test performance. This contrasts with some of the previous studies of the CDT that have indicated significant differences across age groups or significant correlations with age on neuropsychological screening tools (Liu et al., 2011).

We also did not find a relationship between sex and Papadum test performance. Regarding CDT performance for the patient sample, there was an effect of sex on performance. Women were shown to be more significantly impaired in their dementia ratings therefore this may have influenced results. On the CDT (healthy controls), the effect of sex was just above the cut off for significance, with a small-medium effect size. As the females had a slightly lower level of education it was possible that education differences could account for the modest difference in CDT performance between males and females, and it was noted that in a regression model with education and sex as predictors, only education significantly predicted CDT performance. Previous studies have not found an influence of sex on CDT performance (Shanhu et al., 2019; Kim & Chey, 2010).

In relation to diagnostic accuracy; the area under the curve (AUC) measures (which were all in the 'fair' range), sensitivity, and specificity for CDT and both forms of the Papadum test were similar, albeit modest. The fact that a test aimed at testing cognitive skills such as planning, praxis, and visuo-spatial functions was completed relatively successfully by some patients with dementia is not surprising given that these cognitive skills are often not impaired in the mild-moderate stages of dementia, particularly Alzheimer's disease.

The Papadum test could provide a viable alternative as a screening tool to the CDT for use with people who are illiterate or have low levels of education, as results were not influenced by level of education. If a participant performs poorly on the Papadum, it could indicate difficulties with visuo-spatial function and/or with planning and would indicate further investigation is required. Other ways of testing the specific constructs of interest may be required. Thus, as with the CDT, the Papadum test could be used as a screening tool, as a means of detecting problems that can be investigated in further assessment (i.e., favouring sensitivity over specificity). It is most realistic that the paper version of the Papadum would be utilised in clinic settings.

Future research

Future studies should assess those who are illiterate to assess diagnostic accuracy of the Papadum test in relation to the diagnosis of dementia. In this study, the Rouleau scoring system was used but was adapted slightly to include information such as including two scripts (e.g., Bengali and English) in the clock face, which was believed to be important to capture the cultural differences. Future research could also include using various scoring versions of the CDT and comparing to the Papadum/Pizza test. It is

thought that the variability of scoring methods can mean that the results of various studies are difficult to compare (Shulman et al., 2000). Future research could also consider the inclusion of qualitative analysis.

Regarding future validation studies, it will be important to assess the influence that any motor impairment or visual difficulties have on the Papadum test performance.

There was a wide range of education amongst the participants in this study, though the average length of education for both patients and healthy controls was well above the average for the population in India. This in part reflects the fact that only literate participants were included to ensure that participants could do both the CDT and Papadum tasks. It also reflects the population who attended the hospitals from where participants were recruited. It is important that future studies aim to match participant recruitment closer to population averages to increase confidence that results would generalise to the wider population. Recruitment of larger samples would also allow for more precise matching between patients and healthy controls. The use of regression models allowed age, education and sex to be included as covariates when examining the relationship between CDT and Papadum performance, but it would be good for future studies to match samples on these demographic factors.

One limitation of the study was that the data were collected in India and some of the scoring to examine reliability was undertaken using photographs of the stimuli. Although it is believed that this did not compromise the scoring, it would be better in future (where possible) to score original materials (CDT or Papadum) as this may improve reliability further.

CONCLUSION

Education had a significant association with the CDT but not with the Papadum test. Across the whole sample there was a significant correlation between CDT and Papadum but not within separate groups of healthy controls and patients. Diagnostic accuracy for CDT was similar to that for Papadum.

This study further provides evidence to support the findings from the Pizza test study (Manoj et al., 2015), that this is a screening tool that could be used as an alternative to the CDT that is not impacted by education. Although further validation studies are required, evidence from this study would support the use of this task as an alternative to the CDT with individuals who have a lack of education or literacy as a means of assessing planning, visuospatial and praxis skills that may be impaired as part of a neurodegenerative process.

REFERENCES

Ainslie, N.K., & Murden, R.A. (1993) Effects of education on the clock drawing dementia screen in non demented elderly persons. *Journal American Geriatric Society, 41,* 249-252.

Ardila A., Rosselli M., & Rosas P. (1989). Neuropsychological assessment in illiterates: Visuospatial and memory abilities. *Brain and Cognition ,11*(2),147-66.

Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. *Neuropsychology Review, 15*(4), 185–195.

Ardila, A., Bertolucci, P.H., & Braga, L.W. (2010) Illiteracy: The neuropsychology of cognition without reading. *Archives Clinical Neuropsychology, 25,* 689-712.

Critchley, M.(1953). *The parietal lobes*. New York, NY: Hafner Publishing Company.

Crombie, M. & Evans, J. (2021). The Impact Of Education On The Clock Drawing Test (CDT): A Systematic Review. Submitted for publication.

De Noronha, Í.F.C.D., Barreto, S.D.S., & Ortiz, K.Z. (2018). The influence of education on performance of adults on the Clock Drawing Test. *Dementia & Neuropsychologia, 12*(1).

Das, S.K., Banerjee, T.K., Mukherjee, C.S., Bose, P., Hazra, A., Dutt, A., Das, S., Chaudhuri, A., Raut, D.K. (2006). An urban community-based study of cognitive function among non-demented elderly population in India. *Neurology Asia, 11,* 37 – 48.

Duro, D., Tábuas-Pereira, M, Freitas, S., Santiago, B., Boteldo, M.A., & Santana, I. (2018). Validity and Clinical Utility of Different Clock Drawing Test Scoring Systems in Multiple Forms of Dementia. *Journal of Geriatric Psychiatry and Neurology 31*(2).

Ferri, C.P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., & Ganguli, M. (2005) Alzheimer's Disease International. Global prevalence of dementia: A Delphi consensus study. *Lancet, 366*(9503), 2112–2117.

Folstein, M., Folsten, S., & McHugh, P. (1975). Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *Journal Psychiatry Research, 12*,189–198.

Hsieh, S., Schubert, S., Hoon, C., Mioshi, E., & Hodges, J.R. (2013). Validation of the Addenbrooke's Cognitive Examination III in Frontotemporal Dementia and Alzheimer's Disease. *Dementia Geriatric Cognitive Disorders, 36,* 242-250.

Hughes, C.P., Berg, L., Danziger, W.L., Coben, L.A., & Martin, R.L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry, 140,* 566–572.

Kalafat, M., Hugonot-Diener, L., & Poitrenaud, J. (2003). Standardisation et étalonnage français du « Mini Mental State » (MMS) version GRECO [French standardization of the Mini Mental State (MMS), GRECO's version]. *Revue de Neuropsychologie, 13,* 209–236.

Khan, F. (2011). Dementia in India: it's high time to address the need!. *Andhra Pradesh Journal of Psychological Medicine, 12*(2), 64-8.

Kim, H., & Chey, J. (2010). Effects of education, literacy, and dementia on the clock drawing test performance. *Journal of International Neuropsychology Society, 16,* 1138-1146.

Liu, K.P., Kuo, M.C., & Tang, K.C., (2011). Effects of age, education and gender in the Consortium to Establish a Registry for the Alzheimer's Disease (CERAD)-Neuropsychological assessment battery for Cantonese speaking Chinese elders. *International Psychogeriatrics, 23*,1575-1581.

Nielsen, T.R. & Jorgensen, K. (2013). Visuoconstructional abilities in cognitively healthy illiterate Turkish immigrants: A quantitative and qualitative investigation. *The Clinical Neuropsychologist, 27*(4), 681–692.

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C.P. (2013). The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimers & Dementia, 9*(1), 63–75.

Rouleau, I., Salmon, D.P., Butters, N., Kennedy, K.C., McGuire, K.(1992) Quantitative and qualitative analyses of clock drawings in Alzheimer' and Huntington' disease. *Brain and Cognition, 18,*70-87.

Royall, D.R., Cordesa, J.A., & Polka, M.(1998) Clox: an executive clock drawing task. *Journal Neurological Neurosurgical Psychiatry, 64*, 588–594.

Scanlan, J.M., Brush, M., Quijano, C., & Borson, S. (2002) Comparing clock tests for dementia screening: naive judgments vs formal systems-what is optimal*? International Journal Geriatric Psychiatry,17,*14–21.

Shanhu, X., Linhui, C., Xiaoqing, J., Jing, Y., Saizhu, X., Ying, X., Caixia, L., & Yu J. (2019) Effects of age and education on clock-drawing performance by elderly adults in China. *The Clinical Neuropsychologist, 33*(1), 96-105.

Shulman, K. I. (2000). Clock-drawing: Is it the ideal cognitive screening test? *International Journal of Geriatric Psychiatry, 15*, 548-561.

Stern, Y., Gurland, B., Tatemichi, T. K., Tang, M. X., Wilder, D., & Mayeux, R. (1994). Influence of education and occupation on the incidence of Alzheimer's disease. *The Journal of the Medical Association, 271*(13),1004-10.

Manoj, M., Crawford, S., Dutt, A., Ghosh, A., Nandi, R., and Evans, J.J. (2015) Pizzas and Papadums : An alternative to the Clock Drawing Test for the assessment of cognition. Poster presented at the International Neuropsychological Society 2015 Mid-Year Meeting / 5th ASSBI Pacific Rim Conference July 1-4, 2015 Sydney, Australia.

Nielsen, T.R. & Jorgensen, K. (2013). Visuoconstructional abilities in cognitively healthy illiterate Turkish immigrants: A quantitative and qualitative investigation. *The Clinical Neuropsychologist, 27*(4), 681–692.

Tripathi, R. K., Verma, Y., Srivastava, A., Shukla, T. S., Usman, K., Ali, W., & Tiwari, S. C. (2020). Usefulness of clock-drawing test in Indian older adults with diabetes mellitus. *Indian journal of psychiatry, 62*(1), 59–65.

UNESCO Institute of Statistics (2017, September 1). *Literacy Rates Continue to Rise from One Generation to the next*. Retrieved from http://uis.unesco.org/en/topic/literacy

Von Gunten, A., Ostos-Wiechetek, M., Brule, J., Vaudaux-Pisquem, I., Cattin, S., & Duc R. (2008). Clock-drawing test performance in the normal elderly and its dependence on age and education. *European Neurology, 60*, 73-78.

World Health Organization and Alzheimer's Disease International. (2012). Dementia a public health priority.World Health Organization press. Retrieved from:

https://www.who.int/mental_health/publications/dementia_report_2012/en/

APPENDICES

Appendix 1: Guidelines for submission to Clinical Neuropsychologist Journal

INSTRUCTIONS FOR AUTHORS

The Clinical Neuropsychologist (TCN) provides in-depth discussions of matters germane to the practicing clinical neuropsychologist. Clinical neuropsychology is a rapidly expanding field, there is a need for airing of empirical data, models, concepts, and positions pertaining to educational, clinical, and professional issues. *TCN* is designed to provide a forum for such presentation and discussions.

Submission

Manuscripts must be submitted through the journal's Scholar One website, http://mc.manuscriptcentral.com/ntcn. Questions for the editor may be addressed to: Jerry J. Sweet at **jerrysweet@uchicago.edu**

Each manuscript must be accompanied by a statement that it has not been published elsewhere and that it has not been submitted simultaneously for publication in another source. Authors are responsible for obtaining permission to reproduce copyrighted material from other sources and are required to sign an agreement for the transfer of copyright to the publisher. All accepted manuscripts, artwork, and photographs become the property of the publisher. Authors are responsible for disclosing any funding sources and financial interests that could create a potential conflict of interest (see volume 18, page 1). All parts of the manuscript should be typewritten, double-spaced, with margins of at least one inch on all sides. Authors should also supply a shortened version of the title suitable for the running head, not exceeding 50 character spaces. Each article should be summarized in an abstract of not more than 200 words. Avoid abbreviations, diagrams, and reference to the text in the abstract.

Before uploading your final accepted manuscript authors must check the following:

- The first page of the article contains all author(s) contact details.
- The number of tables is the same as the number of table legends, table and figure numbers are consecutive and figures are numbered.
- Journal titles in text or references must not be abbreviated. Please supply full journal names.
- You must supply your accepted manuscript in document format, not as a pdf.

References

Cite in the text by author and date (Smith, Jones, & Brown, 1983). Prepare reference list in accordance with the APA Publication Manual, 6th ed. Examples:

Journal: Adlington, R. L., Laws, K. R., & Gale, T. M. (2009). The Hatfield Image Test (HIT): A new picture test and norms for experimental and clinical use. *Journal of Clinical and Experimental Neuropsychology, 31*, 731–753. doi:10.1080/13803390802488103

Book: Hammen, H., & Watkins, E. (2008). *Depression* (2nd ed.). Hove, UK: Psychology Press.

Chapter in a Book: Youngjohn, J. (2008). Lyme disease: Consideration of malingered disability. In J. E. Morgan & R. O. Gervais (Eds.), *Neuropsychology of malingering casebook* (pp. 254–264). Hove, UK: Psychology Press.

Illustrations

Illustrations submitted (line drawings, halftones, photos, photomicrographs, etc.) should be clean originals or digital files. Digital files are recommended for highest quality reproduction and should follow these guidelines:

- 300 dpi or higher
- Sized to fit on journal page
- JPEG or TIFF format only
- Submitted as separate files, not embedded in text files

Color illustrations will be considered for publication; however, the author will be required to bear the full cost involved in their printing and publication. The charge for the first page with color is $900.00. The next three pages with color are $450.00 each. A custom quote will be provided for color art totaling more than 4 journal pages. Good-quality color prints should be provided in their final size. The publisher has the right to refuse publication of color prints deemed unacceptable.

Tables and Figures

Tables and figures (illustrations) should not be embedded in the text, but should be included as separate sheets or files. A short descriptive title should appear above each table with a clear legend and any footnotes suitably identified below. All units must be included. Figures should be completely labeled, taking into account necessary size reduction. Captions should be typed, double-spaced, on a separate sheet. All original figures should be clearly marked in pencil on the reverse side with the number, author's name, and top edge indicated.

Proofs

Page proofs are sent to the designated author using Taylor & Francis' EProof system. They must be carefully checked and returned within 48 hours of receipt.

Offprints/Reprints

Reprints are available for purchase upon registration with Rightslink, our authorized reprint provider. Authors will need to create a unique account and register with Rightslink for this free service. The link is provided at the time of page proof review.

Grand Rounds in Clinical Neuropsychology

Overview
Grand Rounds in Clinical Neuropsychology is devoted to case presentations of interesting, timely, important, or unusual cases. Cases of interest to be considered may represent unusual presentations of well-known disorders/syndromes, rarely seen disorders, 'classic' or prototypical neuropsychological syndromes (textbook presentations), or other cases of distinction. Adult and child cases will be considered. Criteria for publication include a well-documented history of the patient, medical/neurologic/psychiatric findings, neuroimaging (*preferred, but not required*), neuropsychological evaluation, discussion, and conclusions. Cases should be instructive and focus on the contributions that competent neuropsychological assessment makes in terms of:

(1) elucidating brain-behavior relationships;
(2) determining the functional status of patients; and
(3) instructing intervention, treatment, rehabilitation, education, etc.

TCN Grand Rounds in Clinical Neuropsychology, unlike the aims and scope of *Neurocase*, a sister publication of T&F, will not focus on elucidating theoretical aspects of brain-behavior relations, but instead will focus on well-known and documented aspects of "behavioral geography" particularly as illustrated in neurological or neuropsychiatric conditions.

Format
Two general formats will be utilized:

(1) The traditional case presentation format, where the diagnosis/syndrome/disorder known in advance, usually in the title, and
(2) clinical problem-solving cases, where the diagnosis/syndrome/disorder is not revealed until the conclusion, similar in organization to the Grand Rounds Presentations of the Massachusetts General Hospital in the *New England Journal of Medicine*.

Page Limitations
A maximum of 35 pages, **double spaced**, inclusive of references. Allowances will be made for slight departures, dependent on the case.

Neuropsychological Testing
Neuropsychological assessment may follow a flexible or fixed battery approach, as long as the referral question is appropriately and fully answered. Truncated or focused assessment batteries may be acceptable in specific cases; comprehensive testing is not always indicated or necessarily appropriate. Authors are encouraged to present cases that illustrate assessment skills and clinical knowledge at an advanced level, illustrative of strong clinical judgment skills.

Literature Review
Sufficient review of the literature regarding the syndrome is necessary, but this should not be overly elaborated and detailed. Where presented findings are congruent or at odds with 'classic' cases, this should be so noted and appropriately referenced. Presentations of a relatively 'pure' prototypical disorder should reference the original article (e.g., a circumscribed Gerstmann's syndrome).

Neuroimaging
Authors are encouraged to present CT/MRI when available, documenting the relationship between neuropsychological test findings and cerebral abnormalities. Where definitive "diagnoses" have been made in a neurological or other medical venue, confirmatory evidence must be presented. These may include neuroimaging where available (this is preferred; we will publish b&w CT/MRI), scans, pathology reports, inclusive of examination by board certified medical specialists.

Types of Cases
The editor(s) will consider most types of clinical cases. These may include well-known and documented neurological conditions, low base rate (rarer or unusual) disorders, and common disorders with an unusual presentation or neuropsychological findings (e.g., a large cerebral neoplasm with a paucity of neuropsychological test abnormalities), neurodevelopmental conditions (in the case of a pediatric presentation), and disorders of controversial etiology (e.g., CFS). Cases involving poor effort, frank malingering, factitious disorders and the like must include appropriate use of SVTs and documentation of a lack of medical evidence. Some psychiatric conditions with known or putative CNS abnormalities may be appropriate (e.g., schizophrenia) at the discretion of the editor(s).

Grand Rounds in Clinical Neuropsychology is an exciting new section in *TCN*. As part of one of neuropsychology's leading journal publications, the highest professional standards of practice and publication are expected.

Manuscripts for the Grand Rounds section must be submitted through the journal's Scholar One website, http://msc.manuscriptcentral.com/ntcn and clearly label your submission as intended for consideration in Grand Rounds section.

APPENDICES 1.1: SYSEMATIC REVIEW

Appendix 1.1.1.: Adapted QATSO tool (Wong et al. 2008)

	Design	Confounding variable	Validity of predictor measure	Validity of outcome measure
0	Insufficient reporting of sampling to determine the range of years of education amongst participants	No attempt to control confounding variables during recruitment or analysis	Minimal rating of education level- e.g. literate vs illiterate	No evidence of use of a standardised approach to administration or scoring of the CDT. Unclear who scored the test.
1	Source of participants reported but restricted to a limited educational range – participants from a single level of education (e.g University students)	Some attempt to control for confounding variables (E.g. age or gender)	Level of education (e.g. classification into university, high school etc.) or years of education reported.	Standardised administration and scoring method reported.
2	Sample includes participants from a wide range of years/levels of education e.g. participants from at least two levels of education (e.g. University/High school)	Age, gender, controlled for and any other variables eg. Place of residence (urban vs rural), occupation	Level/years of education, plus a measure of quality of education reported	Trained clinicians/researchers using a standardised administration and scoring method. Inter rater reliability of test

Appendix 1.1.2: Ratings on QATSO for each study

Study name	Design	Confounding variable(s)	Validity of predictor measure	Validity of outcome measure	Total
Ainslie & Murden (1993)	2	0	1	2	**5**
Balduino et al. (2020)	1	1	1	1	**4**
Bozikas et al. (2008)	2	2	1	1	**6**
DeNoronha et al. (2018)	2	1	1	1	**5**
Fabricio et al. (2013)	2	2	1	2	**7**
Kim & Chey (2010)	2	2	1	2	**7**
Caffarra et al. (2011)	2	2	1	1	**6**
Paganini et al. (2001)	1	2	1	2	**6**
Hubbard et al. (2008)	2	2	1	2	**7**
Turcotte et al. (2018)	2	2	1	2	**7**
Leung et al. (2005)	2	1	1	2	**6**
Marcopulos et al.(1997)	1	2	1	1	**5**
Merims et al. (2018)	2	2	1	2	7
Nitrini et al. (2004)	2	2	1	1	**6**
Santana et al. (2013)	2	2	1	2	**7**
Seigershmidt et al. (2002)	2	2	1	2	**7**
Sicialano et akl (2016)	2	2	1	2	**7**
Shanhu et al. (2019)	2	2	1	1	**6**
Shao et al. (2020)	2	2	1	2	**7**
Sugawara et al. (2010)	2	2	1	1	**6**
Von Gunten et al. (2005)	2	2	1	1	**6**

Appendix 1.1.3: Ratings on QATSO for 33% of studies by peer reviewer

STUDY	Design	Confounding variable(s)	Validity of predictor measure	Validity of outcome measure	Total
Bozikas et al. (2008)					
	2	2	1	1	**6**
Kim & Chey (2010)					
	2	2	1	2	**7**
Caffarra et al. (2011)					
	2	2	1	1	**6**
Paganini et al. (2001)					
	1	2	1	2	**6**
Hubbard et al. (2008)					
	2	2	1	2	**7**
Turcotte et al. (2018)					
	2	2	1	2	**7**
Nitrini et al. (2004)					
	2	2	1	1	**6**

APPENDICES 2: MAJOR RESEARCH PROJECT

Appendix 2.1 Ethics approval from Apollo

Apollo Gleneagles HOSPITALS KOLKATA

Organization Accredited by Joint Commission International

INSTITUTIONAL ETHICS COMMITTEE

Chairman
Prof Sukumar Mukherjee

Member Secretary
Dr Jayanta Kr Gupta

Ex Officio Member
Dr Dipankar Ganguly
Director, Medical Services

Members

Dr Saibal Mukherjee

Dr Tirthankar Chaudhury

Dr Lawni Goswami

Dr Enam Murshed Khan

Prof Santanu K Tripathi

Mr Prabir Basu

Mr Indranil Ghosh

Mr Tushar Kanjilal

Ms Alo Sengupta

Mr Debasis Chakrabarti

Mr Sujit Kumar Saha

Ms Tania Das

- Mini Mental Status Examination (MMSE)
- Addenbrooke's Cognitive Examination III
- Clock Drawing Test (CDT) performance sheet
- The Pizza / Papadum test Scoring sheet

B. The following members were present in the IEC meeting, in accordance with the ICH-GCP guidelines, held on 14th December 2013, at 9:00 AM in Apollo Gleneagles Hospital Limited. This satisfies the quorum necessary for such meetings of the IEC.

C. The decision was arrived at through consensus. Neither PI nor any of proposed study team members was present during the decision making of the IEC.

Sr. No	Name	Affiliation	Role in the EC
1	Prof. Sukumar Mukherjee	Calcutta Medical Research Institute	Chairman
2	Dr Tirthankar Chaudhuriy	Apollo Gleneagles Hospitals Limited	Acting Member Secretary
3	Dr Enam Murshed Khan	Apollo Gleneagles Hospitals Limited	Basic Medical Scientist
4	Dr. Saibal Mukherjee	Apollo Gleneagles Cancer Hospital	Clinician
5	Dr Jayanta Kr Gupta	Apollo Gleneagles Hospitals Limited	Member secretary
6	Mr Debasis Chakrabarti	Compassionate Crusaders Trust	Social worker
7	Ms. Alo Sengupta	Apollo Gleneagles Hospitals Limited	Lady Representative
8	Mr. Prabir Basu	Supreme Court	Lawyer
9	Mr. Indranil Ghosh	Apollo Gleneagles Hospitals Limited	Lay person
10	Mr Sujit Kumar Saha	Apollo Gleneagles Heart Centre	Lay Person
11	Dr Dipankar Ganguly	Apollo Gleneagles Hospitals Limited	Ex Officio Member

D. Continued approval is conditional upon your compliance with the following requirements:

a. In the event of any protocol amendment, IEC must be informed. The exact alteration/amendment should be specified and

Hospital : 58, Canal Circular Road, Kolkata-700 054, Tel. : 91 33 2320 3040/2320 2122, Fax : 91 33 2320 5184/2320 5218
Gariahat Clinic : 48/1F, Leela Roy Sarani, (Gariahat), Kolkata-700 019, Tel. : 91 33 2461 8028/8079/8451/8547/9482/9483, Fax : 91 33 2461 8180
E-mail : hospital@apollogleneagles.in CIN - U33112WB1988PLC045223

Emergency Service Dial 1066

Apollo Gleneagles HOSPITALS KOLKATA

Organization Accredited by Joint Commission International

INSTITUTIONAL ETHICS COMMITTEE

IEC Ref:IEC/2013/12/40 **Approval Date:** 24th December 2013

Dr Aparna Dutt
Department of Neurology
Principle Investigator
Senior Consultant
Apollo Gleneagles Hospitals Kolkata

Dear Dr Dutt,

Sub: Institutional Ethics Committee (IEC) Approval of New Application

Protocol Title: Is the Pizza / Papadum test a Good Alternate to the Clock Drawing Test?

This is with reference to your application dated 10th December 2013 vide which the study documents for the referenced study were submitted to the Institutional Ethics Committee for review and approval.

We are pleased to inform you that the IEC has approved the above research project to be conducted in Apollo Gleneagles Hospitals, Kolkata in its current form.

The reference number for this study is **IEC/2013/12/40.** Please use this reference number for all future correspondence.

A. In the meeting held on 14th December 2013 at 9.00 AM in Apollo Gleneagles Hospitals, Kolkata, the IEC conducted a scientific and ethical review of the above study. The following documents were reviewed and approved:

- Research Protocol; Version 1, Date: 10/12/13
- Informed Consent form for Healthy Volunteers; Version 1, Date: 10/12/13
- Informed Consent Form for Patient, Version 1, Date: 10/12/13
- Demographic sheet- for Healthy volunteers and patients

Hospital : 58, Canal Circular Road, Kolkata-700 054, Tel. : 91 33 2320 3040/2320 2122, Fax : 91 33 2320 5184/2320 5218
Gariahat Clinic : 48/1F, Leela Roy Sarani, (Gariahat), Kolkata-700 019 Tel : 91 33 2461 8028/8079/8451/8547/9482/9483, Fax : 91 33 2461 8180
E-mail : hospital@apollogleneagles.in CIN - U33112WB1988PLC045223

Emergency Service Dial 1066

Apollo Gleneagles HOSPITALS

INSTITUTIONAL ETHICS COMMITTEE

Organization Accredited by Joint Commission International

indicated where the amendment occurred in the original project. (Page no. Clause no. etc.)

b. No deviation from, or change to, the protocol approved by the Institutional Ethics Committee should be implemented without the prior written approval of the Institutional Ethics Committee. Deviations/ changes to the approved protocol may be implemented without prior approval of the Institutional Ethics Committee only when necessary to eliminate immediate hazards to subjects

c. c. Any deviation from, or a change of, the protocol to eliminate an immediate hazard should be promptly reported to the IEC within seven calendar days.

d. Please submit the following to the IEC

a. Yearly progress of the study.
b. All SAEs including local deaths whether related or not should be reported to IEC immediately within 24 hours of occurrence.
c. Report(s) on any new information that may adversely affect the safety of the subject or the conduct of the study.
d. Study completion report is to be submitted within 4 weeks of study completion or termination.

F. It is mandatory to register all clinical trials with www.ctri.in. You are strongly encouraged, in consultation with your sponsor (where applicable), to register clinical trial with http://www.clinicaltrials.gov. The registration is free.

With best wishes,

Yours Sincerely,

Dr Jayanta Kumar Gupta
Member Secretary
Institutional Ethics Committee (IEC)
Apollo Gleneagles Hospital, Kolkata

Dr. Jayanta Kumar Gupta
Member Secretary,
Institutional Ethics Committee
Apollo Gleneagles Hospitals, Kolkata

[redacted]

Appendix 2.2 Ethics approval from Duttanager

DUTTANAGAR MENTAL HEALTH CENTRE DMHC

INSTITUTIONAL RESEARCH ETHICS COMMITTEE

To
Dr Aparna Dutt
Consultant Neuropsychologist
Neuropsychology & Clinical Psychology Unit
Duttanagar Mental Health Centre
Kolkata 700 077

Dear Dr Aparna Dutt,

The Institutional Research Ethics Committee (IREC) of Duttanagar Mental Health Centre has reviewed and discussed your application to conduct the study entitled **"Is the Pizza / Papadum test a Good Alternate to the Clock Drawing Test?"** at Duttanagar Mental Health Centre, Kolkata.

The IREC conducted a scientific and ethical review of the following documents in the meeting held on 12th June 2015 at 12pm in Duttanagar Mental Health Centre, Kolkata.

i. Research Protocol
ii. Informed Consent Form (ICF): Healthy Volunteer
iii. Informed Consent Form (ICF): Patient
iv. Demographic Information Sheet: Healthy Volunteer & Patient - Illiterate
v. Demographic Information Sheet: Healthy Volunteer & Patient - Literate
vi. Mini Mental State Examination (MMSE) / Bengali Mental State Examination / (BMSE)
vii. Addenbrooke's Cognitive Examination III (ACE-III) – Bengali Version
viii. The Pizza / Papadum Test
ix. Hospital Anxiety and Depression Scale (HADS)
x. Geriatric Depression Scale (GDS)

We are pleased to inform you that the full ethical approval has been granted to your study.

Approval no	**DMHC/IREC/02/2015**
Approval date	12 June 2015
Expiry date	12 June 2018
IREC Decision	**Approved**

DUTTANAGAR MENTAL HEALTH CENTRE, DUTTANAGAR, KOLKATA – 700 077
Phone: (033) 25572482/65052244 Email: dmhc@vsnl.com web: www.dmhc.co.in

The following members of the ethics committee were present at the IREC meeting

Srl No.	Name	Designation	Role in the EC
1	Dr Anil Bhuson Dutt	Consultant Psychiatrist	Chairperson
2	Mr Namit Biswas	Retd Spl. Secretary, Family Welfare, West Bengal	Member Secretary
3	Dr Kishan Pradhan	General Secretary Calcutta Heart Clinic & Hospitals, Kolkata	Clinician
4	Dr Dharitri Dutt	Physician National Institute of Cholera and Enteric Diseases	Clinician
5	Mr Sumanta Chaudhury	Retired Colonel, Indian Army	Ethicist
6	Dr Arindam Biswas	Project Investigator & Research Co-ordinator S.N. Pradhan Centre for Neuroscience University of Calcutta, Kolkata	Medical Scientist
8	Mr Y M Nandy	Lawyer, Calcutta High Court	Legal expert
9	Mrs Rati Basu	School Teacher, Shantiniketan, West Bengal	Lay Person

Please follow the standard conditions:

a. conduct the project strictly in accordance with the proposal submitted and granted ethics approval, including any amendments made to the proposal required by the IREC
b. report the IREC about any change in the protocol
c. provide a progress report to the IREC annually
d. provide a 'final report' after completion of the project
e. advise in writing if the project has been discontinued.

ETHICS COMMITTEE APPROVED Duttanagar, Mental Health Centre

You may now commence your project.

With Best Wishes

Yours Sincerely,

Mr. Namit Biswas
Member Secretary
Institutional Research Ethics Committee
Duttanagar Mental Health Centre
Kolkata

Mr Namit Biswas
Member Secretary
Institutional Research Ethics Committee (IREC)
Duttanagar Mental Health Centre, Kolkata

Appendix 2.3: Rouleau method for CDT scoring (adapted – see italics)

Integrity of the clockface (maximum: 2 points)
2: Present without gross distortion
1: Incomplete or some distortion
0: Absent or totally inappropriate

NOTES-
1.The clockface is only the circle of the clock.
2.While scoring for distortion, keep in mind that it should be roughly circular.
3. An overlap in the circle as shown in the picture should not be considered as a distortion.

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 arrangement. *Numbers written in two or more scripts*.
2: Numbers missing or added but no gross distortions of the remaining numbers.
Numbers placed in counterclockwise direction. Numbers all present but gross distortion in spatial layout (i.e., hemineglect, numbers outside the clock)
1: Missing or added numbers and gross spatial distortions
0: Absence or poor representation of numbers (*include illegible numbers)*

NOTES -
1. Poor representation of numbers (score 0) also includes illegible numbers.
2. The criteria for a score of 3 will also include numbers written in two or more scripts.

Presence and placement of the hands (maximum: 4 points)
4: Hands are in 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

NOTES –
The difference between 'slight errors in placement of hands' and 'major errors in the placement of hands' –
1. a. slight errors in the placement of hands – hands drawn away from the current number but still close to it.
b. major errors in the placement of hands – hands pointed at the wrong number.
2. Do not deduct any points if the hands do not meet at the centre.

Appendix 2.4 – Scoring criteria for Papadum Test

Scoring criteria :

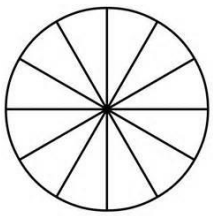
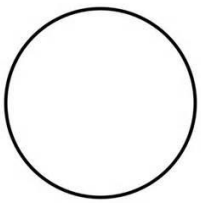
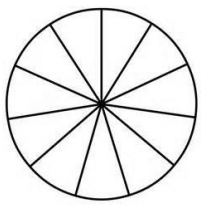
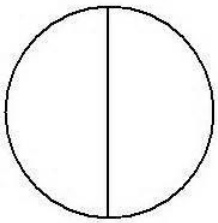
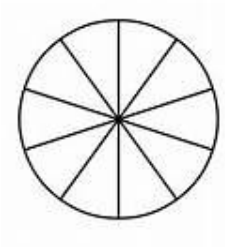
The Papadum test was scored based on the following three criteria:

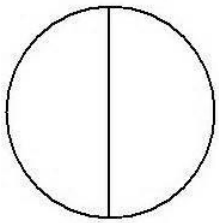
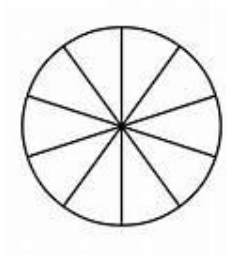
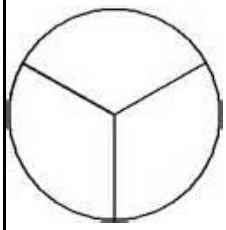
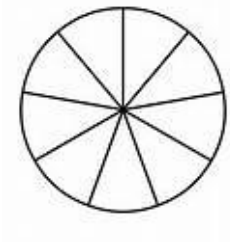
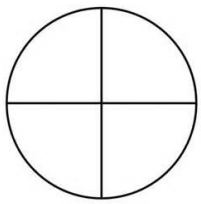
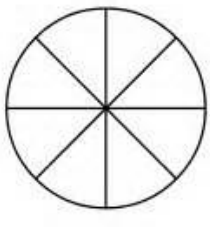
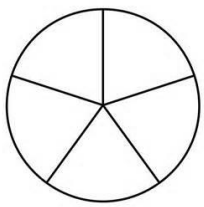
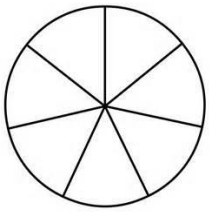
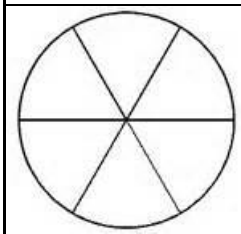
1. Number of pieces 2. Shape of the pieces 3. Size of the pieces.

The maximum total score for the Papadum test is 18.

1. Number of pieces (note that for this criteria shape/size of the pieces does not matter)

1 point is awarded for each piece, up to a maximum of 6. If more than 6 pieces are produced points are deducted for each extra piece over 6 (e.g. 2 pieces = 2 points; 3 pieces =3 points; 6 pieces = 6 points; 7 pieces = 5 points; 10 pieces = 2 points; 12 pieces = 0 points). MAX SCORE = 6

NUMBER OF PIECES	**SCORE**
 = 12  = 0	0
 = 11	1
 =2  =10	2

 =2  =10	2
 =3  =9	3
 =4  =8	4
 =5  =7	5
 =6	6

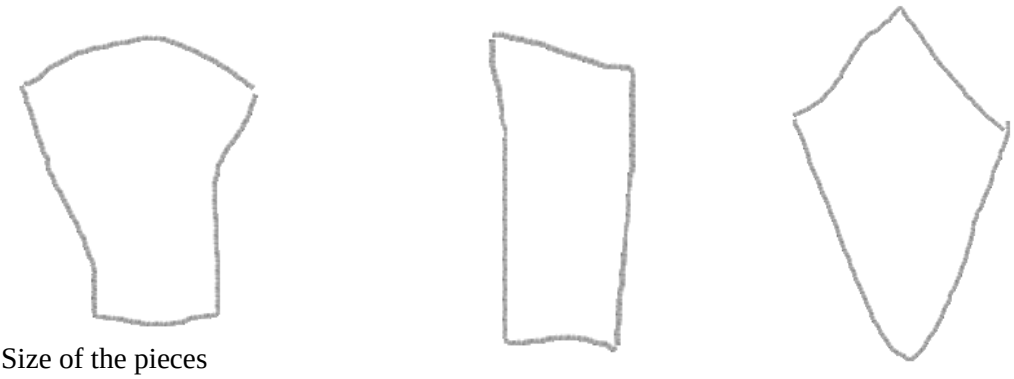
2. Shape of the slices

1 point is awarded for each piece that is triangular in shape (i.e. has three sides) MAX SCORE = 6.
NOTE: If there are more than six pieces, the score is the number of pieces that are triangular up to a

maximum of six (e.g. if there are eight pieces, six of which are triangular and two are square, score is six. However if there are eight pieces, four triangular and four square, the score is four).

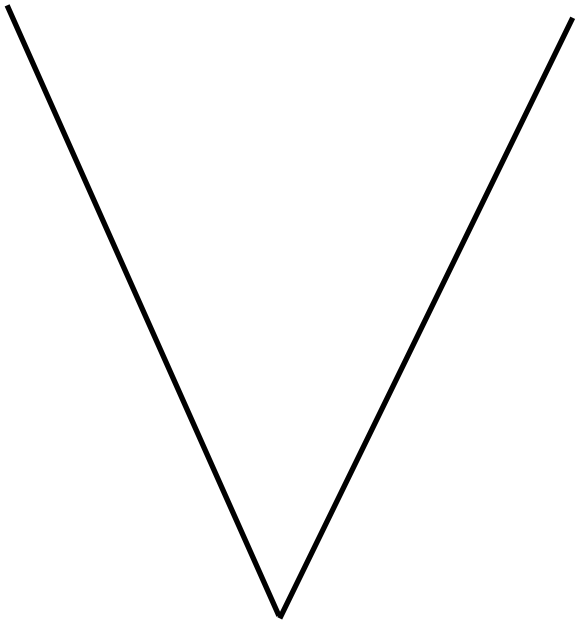
1 point examples

0 point examples

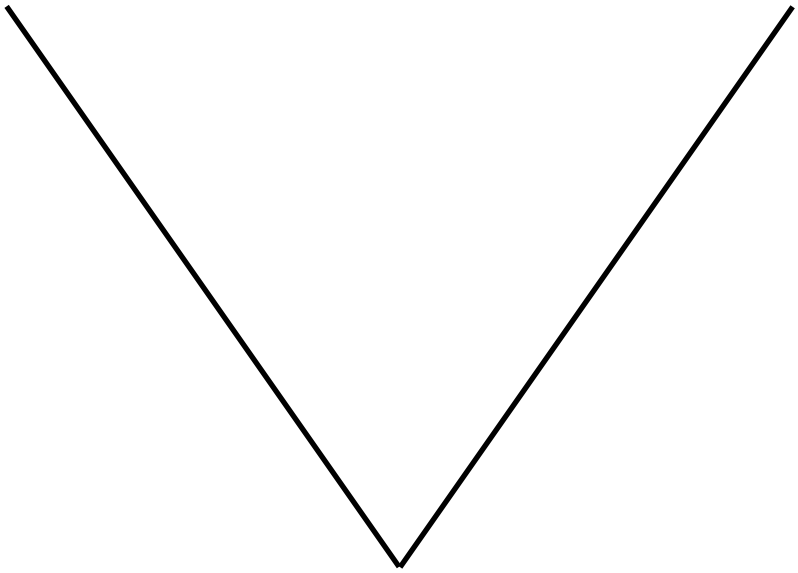


3.Size of the pieces

For pieces that are not triangular in shape, score 0 points. For pieces that are triangular, for each piece that is greater than 50 degrees and less than 70 degrees (see template below) score 1 point. MAX SCORE = 6. NOTE: For the 50 degrees template the slice must cover the template to be scored correctly and for the 70 degrees template the slice must be clearly within the boundary of the template to score the point. If there are more than six pieces, the score is the number of pieces that are within the size boundaries up to a maximum of six (e.g. if there are eight pieces, six of which within the size boundaries and two are outside the boundaries, score is six. However if there are eight pieces, four within the size boundaries and four outside the boundaries, the score is four).



50 Degrees



70 Degrees

Appendix 2.5 New MRP Proposal

Student number :

Title

Validation of the 'Papadum Test': an alternative to the Clock Drawing Test for people with low levels of education/literacy.

Abstract

The prevalence of dementia is growing worldwide, and this is particularly the case in developing countries where improvements in life expectancy are currently greater than in developed countries. There is a need for neuropsychological assessment tools and screening tools that are effective in identifying cognitive deficits to allow for early diagnosis and intervention. Levels of education and literacy remain low in many developing countries. Research on various screening tools has shown good accuracy for dementia recognition, however a paucity of research exists regarding validity when using these tools with low educated or illiterate populations.

In several screening assessment tools, such as the Addenbrooke's Cognitive Examination III (ACE III), a clock drawing test (CDT) is used to screen for cognitive deficits. However, CDTs require a certain level of education and familiarity with writing numbers. The Pizza/Papadum test was designed to be sensitive to the same cognitive functions as the CDT, but without any requirement for familiarity with numbers. Recent research on a UK participant sample found that performance on the Pizza test correlated well with performance on the CDT.

The purpose of this present study is to further this research and test whether the "Papadum Test" (culturally adapted to Indian participants) shows similar results. The present study is a correlation study. 117 healthy adults and 56 adults with a diagnosis of dementia, all over the age of 40 were recruited in Kolkata in India. Participants undertook the ACE III and the Papadum test (both a paper and an actual Papadum version). The level of correlation between performance on the Papadum Test (paper and actual version) and the ACE III Clock Drawing Test will be examined. If there is a significant correlation between the Papadum Test and the CDT, this will suggest that the Papadum test would be a viable alternative for inclusion in adapted forms of the ACE III for use with people who are illiterate or have low levels of education.

Introduction and rationale for project

An increase in life expectancy is occurring across the world, with the greatest increase being seen in low/middle income countries (WHO, 2012). Additionally, there is expected to be an increase in the

number of people living with age related diseases, including dementia (Ferri et al., 2005; Prince et al., 2013). It is estimated that by 2040, 71% of 81.1 million dementia cases will be in the developing world (Ferri, 2005). Dementia is considered as one of the leading causes of disability among older people and typically contributes to many systemic and socioeconomic difficulties (WHO, 2012). There is a clear need for effective neuropsychological testing that can be used to help screen for dementias to allow for early diagnosis and intervention within developing populations.

The Clock Drawing Test (CDT) was devised in the 20th century and has been widely used in research protocols and in clinic for screening individuals suspected of having cognitive impairment or dementia (Royall et al 1998, Crichley 1953). In the last thirty years, an evidence base has formed for its use as an early screening tool of cognitive impairment, especially in Alzheimer's disease (Scanlan et al., 2002). It makes demands on a range of cognitive domains; memory, planning, motor skills, visuospatial abilities, visual memory and reconstruction, attention, concentration and auditory comprehension (Royall et al., 1998). It is a brief tool, that has been widely applied internationally and involves asking the patient to draw the face of a clock and then to draw the hands to indicate a particular time. The CDT is included as a subtest in the Addenbrooke's Cognitive Examination (ACE III; Hsieh et al., 2013), which is a widely used cognitive screening tool. The ACE III is a relatively quick and easy to administer measure that assesses five cognitive domains; attention, language, memory, visuospatial and perceptual abilities.

Although the CDT is a very popular screening tool, it was, along with most neuropsychological tests, developed to be used in Western, educated populations. One of the challenges in screening for dementia in developing countries is related to performance variances due to educational and cultural influences. It is widely documented that neuropsychological test performance is considerably influenced by variables such as language, education, and literacy (Ardila et al. (1989); Ardila (2005)). Strong evidence has further developed over the last decade that indicates that CDT results can be influenced by education level (Kim and Chey, 2010; Nielsen & Jorgensen, 2013) which can lower the specificity of the test. The sensitivity and specificity of the CDT test increases with an increase in the years of education (Von Gunten et al., 2008). Several studies have reported significant limitations in the use of the CDT in clinical practice with people with low low levels of education.

De Noronha (2018) found that CDT performance was greatly impacted for illiterate individuals. Illiteracy refers to the inability to read or write a simple message (UNESCO, 2017). Literacy rates in developing countries among older generations and amongst individuals in rural areas who often receive limited or no formal education. Nielsen and Jorgensen (2013) found that healthy illiterate individuals may experience problems with graphomotor construction tasks when asked to engage in the CDT task. As well as the CDT, another commonly used screening tool, the Mini-Mental State Examination (MMSE) has shown to be influenced by schooling, also displaying a lack of validity in

illiterate populations (Kalafat, Hugonot-Diener, & Poitrenaud, 2003). As well as potentially having an influence on performance, low education and illiteracy are considered major risk factors for developing Alzheimer's disease (Stern et al., 1994). Research displays a crucial need for adequate, adapted tools for the neuropsychological testing of illiterate and low-educated patients, which is prevalent in developing regions.

Levels of literacy in India are particularly low - although the country has made significant progress in improving literacy over the years, it continues to be home to 313 million illiterate people (UNESCO, 2017). Tripathi et al. (2020) recently examined the usefulness of CDT in screening Indian older adults for cognitive impairment and found that education and language are significant variables that correlate with CDT performance. Apart from difficulties with screening tools, other difficulties can exist in regard to a general lack of awareness of dementia in the general population in India, in particular attributing common symptoms to 'normal aging' (Khan, 2011).

The proposed study will examine the validity of a task that has been designed to examine similar cognitive skills to the CDT, but which does not require the ability to read or write and therefore may serve as an alternative to the CDT for illiterate or low educated populations. This test can be adapted culturally and was designed to be an educationally unbiased alternative to the CDT. A study carried out in NHS Greater Glasgow and Clyde, Scotland previously used a similar task to the task included in this study, known as "the Pizza test" (Swarna et al., 2015). This task involved the person being provided with a paper circle and being asked to imagine it was a pizza. They were told that they had to divide the pizza between six people equally, so they had to fold and divide the paper into six equal pieces. The test is considered to assess cognitive abilities such as attention, planning, problem solving, visuo-spatial and praxis skills. Results from this provided evidence that this task captures similar cognitive domains to the CDT. No effects of education were observed suggesting that the task may be appropriate for those with low, or no, education. Results also showed that the Pizza test has the ability to distinguish between patients with dementia and those without dementia with an overall good diagnostic accuracy.

The data for this present study were collected by colleagues in Kolkata, India who developed a Bengali version of the ACE III. The proposed alternative test is known as the "Papadum test" as it has been adapted to an Indian population.

Aim

The purpose of the study is to examine the validity of the Papadum Test as an alternative for the clock drawing test (subtest of the ACE III). This study will further examine the influence years of education has on performance.

Hypothesis

Performance on the Papadum test will be significantly correlated with the performance on the CDT.

Method

Participants

A total of 56 literate patients clinically diagnosed with Alzheimer's disease (AD) and Vascular dementia (VaD) participated in the study. Patients with AD and VaD in the mild and moderate stages of the illness as evident from their scores on the Mini Mental State Examination (MMSE) (Folstein et al., 1975)/ Bengali Mental State Examination (BMSE) (Das et al., 2006) as well as their Clinical Dementia Rating scores were invited to participate in the study. The patients were seen at Duttanagar Mental Health Centre, Kolkata, India. Each patient was assessed by a neurologist or a psychiatrist and underwent an MRI or CT scan. Each patient underwent a comprehensive neuropsychological examination by a qualified neuropsychologist which included a range of neuropsychological tests. Patients with AD with a history of cerebrovascular disease or significant changes in the brain suggestive of cerebrovascular pathology or, patients with severe Parkinson's or stroke dementia patients were excluded from the study. None of the patients included in the study had other neurological illnesses, history of psychiatric illness, head injury, major medical illness and substance abuse.

A total of 117 cognitively healthy literate adults individuals were also included in this study who all resided in Kolkata and are in the age range of 40 and above. These participants were a) relatives or friends of patients attending the Neuropsychology and Clinical Psychology unit at Duttanagar Mental Health Centre, b) family members of other patients attending the hospitals, c) volunteer hospital staff, or d) members in the community (acquaintances of the relatives who volunteered for the study).

Inclusion criteria included having a minimum of 1 year of education with a MMSE score of above 25.

The Hospital Anxiety and Depression Scale (HADS) was also administered. Participants with scores below the cut-off point were included.

Exclusion criteria were applied; individuals with cognitive complaints, hearing or vision problems or any history of neurological or psychiatric illnesses were excluded from this study.

The primary language of all participants was Bengali. Participants were selected through selective sampling. Demographic details of age, sex and years of education was gathered by structured interview for each participant.

Ethical approval

Ethical approval for the study was obtained from the institutional ethics committee of the hospital in Kolkata. Informed consent was sought and provided by all participants

Materials

Addenbrooke's Cognitive Examination (ACE) III – Bengali Version

The ACE III – Bengali Version was administered to all participants. The ACE III is a brief cognitive screening tool that assesses five cognitive domains: attention, memory, verbal fluency, language and visuospatial abilities. The maximum total score is 100. The ACE III includes the Clock Drawing Test as a subtest to assess visuospatial abilities.

Participants were asked to draw a clock with all of the numbers and set the hands at 10 after 5.

The clocks were scored quantitatively according to two scoring systems based on the clockface, numbers and hands using the ACE III, 5-point scoring system and also the Rouleau, 10-point scoring system (Rouleau et al., 1992).

The clocks were also scored qualitatively according to the six error types employed by Rouleau et al., (1992). Some of the scoring criteria were modified while using the Rouleau quantitative and qualitative scoring system. Each clock was scored independently by two raters. See appendix 1 for scoring criteria.

Papadum Test

The Papadum test was administered to all the participants. Two versions of this test were used –

a) Paper version: The individual was given a circular piece of paper measuring 18cm in diameter. The instructions were: "Imagine that this is a papadum which you have to share amongst six members in your family. Could you kindly show me how you will tear the paper so that the six members in your family get an equal share".

b) Papadum version: A dried / unfried papadum was given to each participant. They are asked to divide it into six equal 'slices'. The instructions were: "Imagine that you have to share this papadum amongst six members in your family. Could you kindly show me how you will tear the papadum so that the six members in your family get an equal share".

See appendix 1 for scoring criteria.

Analysis

The planned analysis will examine whether there is a correlation between scores on the Papadum Test with the Clock Drawing Test scores. If the data has a linear relationship and is normally distributed then the parametric Pearson product moment correlation will be computed or if the data exhibits non-linear relationship, or non-normally distributed then the non-parametric, Spearman's correlation will be carried out. If the Papadum test is a good alternative to the CDT, it should show a high correlation with the CDT. The effects of age, education and gender on performance on both the CDT and Papadum tests will be examined using regression models to determine, which, if any, demographic factors are associated with test performance. With a total of 173 participants, the correlation analysis will have 80% power to detect correlation as small as r=0.211.

Finally, the diagnostic accuracy of the test will then be examined using Receiver Operating Characteristic (ROC) curves for the patient-control matched samples. Analysis of the ROC curves will be used to determine the sensitivity and specificity of different cut- off scores, including the cut-off of 13/18 previously found to show maximum Youden index (Swarna et al, 2015).

Practical applications

If the findings of this study support the use of this tool as a valid alternative to the CDT, this will further provide evidence to support the findings from the Pizza test. Further validation studies would be required, however evidence from this study would support the use of this task as an alternative to the CDT with individuals who have a lack of education or literacy to assess similar cognitive skills and deficits.

REFERENCES

Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. Neuropsychology Review, 15(4), 185–195.

Critchley M.(1953) The parietal lobes. New York, NY: Hafner Publishing Company.

Ferri, C.P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., & Ganguli, M. (2005) Alzheimer's Disease International (2005). Global prevalence of dementia: A Delphi consensus study. Lancet. 366(9503): 2112–2117.

Hsieh, S., Schubert, S., Hoon, C., Mioshi, E., & Hodges, J.R. (2013): Validation of the Addenbrooke's Cognitive Examination III in Frontotemporal Dementia and Alzheimer's Disease. Dement Geriatr Cogn Disord;36:242-250.

Kalafat, M., Hugonot-Diener, L., & Poitrenaud, J. (2003). Standardisation et étalonnage français du « Mini Mental State » (MMS) version GRECO [French standardization of the Mini Mental State (MMS), GRECO's version]. Revue de Neuropsychologie, 13, 209–236.

Khan, F. (2011) Dementia in India: it's high time to address the need! AP J Psychol Med. 2011; 12(2): 64-8.

Kim, H., & Chey, J. (2010) Effects of education, literacy, and dementia on the clock drawing test performance. J Int Neuropsychol Soc.16:1138-1146.

Nielsen, T.R. & Jorgensen, K. (2013). Visuoconstructional abilities in cognitively healthy illiterate Turkish immigrants: A quantitative and qualitative investigation. The Clinical Neuropsychologist, 27(4), 681–692.

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C.P. (2013). The global prevalence of dementia: A systematic review and metaanalysis. Alzheimers & Dementia, 9(1), 63–75.

Rouleau, I., Salmon, D.P., Butters, N., Kennedy, K.C., McGuire, K.(1992) Quantitative and qualitative analyses of clock drawings in Alzheimer' and Huntington' disease. Brain Cogn.18:70-87.

Royall, D.R., Cordesa, J.A., & Polka, M.(1998) Clox: an executive clock drawing task. Journal Neurological Neurosurgical Psychiatry.64:588–594.

Scanlan, J.M., Brush, M., Quijano, C., Borson, S. (2002) Comparing clock tests for dementia screening: naive judgments vs formal systems-what is optimal? Int Journal Geriatr Psychiatry.17:14–21.

Stern, Y., Gurland, B., Tatemichi, T. K., Tang, M. X., Wilder, D., & Mayeux, R. (1994). Influence of education and occupation on the incidence of Alzheimer's disease. The Journal of the Medical Association, 271(13):1004-10.

Storey, J.E., Rowland, J.T.J., Basic, D., & Conforti, D.A. (2001). A comparison of five clock scoring methods using ROC (receiver operating characteristic) curve analysis. International Journal of Geriatric Psychiatry,16, 394–399.

Tripathi, R. K., Verma, Y., Srivastava, A., Shukla, T. S., Usman, K., Ali, W., & Tiwari, S. C. (2020). Usefulness of clock-drawing test in Indian older adults with diabetes mellitus. Indian journal of psychiatry, 62(1), 59–65.

UNESCO Institute of Statistics (2017, September 1). Literacy Rates Continue to Rise from One Generation to the next. Retrieved from http://uis.unesco.org/en/topic/literacy

World Health Organization and Alzheimer's Disease International. (2012). Dementia a public health priority. World Health Organization press.

Appendix 2.6 Plain English Summary for Original MRP

Student ID:

Title: Comparison of embedded effort testing tools and a standalone effort test in a dementia population

Background

Neuropsychological tests are designed to assess brain function and are used in clinical settings to help with the diagnosis of deficits. A person's performance on these tests can be influenced by different factors, for example their motivation, depression or stress. If any of these factors impact performance, then the results of these tests are not a valid or reliable representation of their ability. Tests have been developed to help to distinguish whether these factors are influencing a person's ability to perform to their best ability. This type of testing is known as effort testing. Little research exists in the area of dementia and this type of testing, however. The British Psychological Society (BPS) recommend that all clinicians should include effort testing as part of their routine assessment, however it has been recognised that many clinicians do not include these tests, for various reasons, such as, not having time. Two types of effort test exist; one type is included/ embedded amongst other tests whilst the other one is a separate test. A separate test known as the Word Memory Test has shown to be useful in this population, however, separate tests can take more time and can also make it more obvious what the clinician is trying to test, therefore this could also impact a person's results. As it is recommended that we use two of these tests, it is important that we establish which other tests are useful with people with dementia.

Two new parts of a test (RBANS)that is already used to assess for neuropsychological impairment have been established and have shown to be effective in looking at performance a non dementia population. Research has not been carried out in a dementia population yet.

Aims

To investigate the two new parts of the RBANS test and compare it to a "gold standard" (WMT) test of effort/performance in those being assessed for dementia.

Consent

Informed consent will be given by participants. Participant information sheets will be provided.

Participants

Will be older adults aged 65 plus who attend for neuropsychological testing in our NHS board to assess for dementia.

Methods

Participants will be recruited who attend for neuropsychological testing to determine whether they meet diagnosis for dementia or not. The purpose of the study will be explained and it will also be explained that it does not require any additional time from the participants as effort testing should be part of routine assessment anyway. The tests will be carried out as part of routine assessment by myself or a Clinical Psychologist.

Ethical issues

Confidentiality will be discussed in interview and data will be kept secure.

Practical applications

This study aims to add to the limited evidence base regarding effort testing in dementia assessments. This may help to inform new parts of tests that could easily be included in testing. The aim is to ensure that we can assess adequately if any factors influence a person's performance during neuropsychological testing meaning they are not performing to their best ability.

References

BPS – British Psychological Society (2009). Assessment of effort in clinical testing of cognitive functioning for adults. Leicester, British Psychological Society

McGuire, C., Crawford, S. and Evans, J. J. (2019) Effort testing in dementia assessment: a systematic review. Archives of Clinical Neuropsychology, 34(1), pp. 114-131.

Appendix 2.7 MRP Original proposal

MRP Proposal

Comparison of embedded effort testing tools and a standalone effort test in a dementia population

Student ID:

Word count: 3294

Title

Comparison of embedded effort testing tools and a standalone effort test in a dementia population.

Abstract

Neuropsychological test data is only valid when appropriate effort has been applied. It is recommended that all neuropsychological assessments should include a standalone and at least one embedded effort test, however, this does not always occur in practice. A lack of knowledge also exists on which tests effectively measure effort in people with dementia. This study aims to further research by evaluating the use of two novel RBANS indices compared against performance on Green's Word Memory Test (WMT) in individuals undergoing assessment for dementia. The WMT is currently the most appropriate test of effort in this population. If the results of the novel RBANS indices are consistent with the WMT results and show good specificity and sensitivity, the incorporation of these indices as embedded tests will be suggested as they are relatively quick and easy to administer, potentially increasing the likelihood that effort testing is routinely practiced.

Introduction and rationale for project

The interpretation of neuropsychological tests depends upon comparison of the examinee's performance with a normative sample and an assumption that the examinee has applied appropriate effort. If an examinee does not perform at a level that reflects their actual abilities in the domain of cognition being assessed, then the assessment will not be valid (Green et al., 2002).

Possible reasons for sub-optimal effort include deliberate underperformance for gain; however, it is important to note that malingering is only one of several reasons for suboptimal performance (Strauss et al., 2006, p.94). Although many studies have focused their investigation of suspect effort in the context of litigation (e.g., Boone et al., (1995); Gervais et al., (2004)), reasons unrelated to financial gain exist that could result in suboptimal effort; such as, depression, stress, fatigue or inadequate investment, lack of interest or cooperation in the testing process.

The British Psychological Society (BPS, 2009) and the American Academy of Clinical Neuropsychology (AACN, 2009) produced effort testing guidelines recommending effort testing measures be incorporated into routine clinical practice.

Two types of effort testing exist; 'embedded' tests are included in a test that has been designed for another purpose. They are generally less time consuming and may be less sensitive to a client perceiving when a measure is assessing for consistency or exaggerated responses (Miele et al., 2012). 'Standalone' effort measures are designed explicitly to test for effort and are used more commonly in practice; however, they usually do not provide any additional independent information on neuropsychological status (Welsh et al., 2012).

Guidelines suggest using a standalone measure and at least one embedded effort test. In practice, this does not always occur routinely, perhaps due to a perceived lack of time or due to clinicians' reliance on clinical judgement (McCarter et al., 2009). It has been found, however, that clinical judgement is no better than chance in testing for effort (Faust et al., 1988).

Assessing suboptimal effort in a dementia population remains problematic. As neuropsychological impairment may be severe, patients might fail effort measures despite putting forth adequate effort (Teichner & Wagner, 2007). Patients with dementia are infrequently included in samples used for effort test validation (Dean et al., 2008). When included, they frequently score below suggested cut-offs for effort. Standalone measures of effort most frequently measure a single domain of function, usually memory, which can pose the risk of confounding the results of response validity testing with severe cognitive deficits in dementia (Teichner & Wagner, 2004). Many studies have excluded patients with dementia in part because of their generally lowered specificity rates. Base rates of malingering are found to be very low, with as few as 2% making claim alleging dementia (Mittenberg et al., 2002), though Mittenberg et al. (2002) did not consider other reasons for lack of effort, for example depression. It is thus unclear whether many effort measures can be reliably used within dementia populations.

Despite the challenges, it is highly important to assess effort during dementia assessments to ensure valid results. Dean et al. (2008) investigated the efficacy of effort indexes in patients with dementia. Most of the effort tests that were investigated had high false-positive error rates, and several of the established cut-offs were inaccurate in assessing for effort in the dementia sample. One of the standalone effort measures investigated was the Test of Memory Malingering (TOMM; Tombaugh, 1996), which aims to discriminate between genuine and feigned memory difficulties. However, across studies, the specificity of the TOMM in people with dementia has ranged from 82% (Greve et al., 2006) to a low of 24% (Teichner & Wagner, 2004). These generally lowered sensitivity and specificity levels suggest that the TOMM is an ineffective effort test for individuals being assessed for dementia.

Caution is therefore needed in interpreting effort measure performance in people with dementia as despite best effort, patients with dementia may fail effort measures, potentially leading to misdiagnosis (Bortnik et al., 2013).

While the performance of many effort tests has been extensively studied among traumatic brain injury populations, relatively few studies have examined the test characteristics of effort tests when used among samples of dementia patients (Dean et al., 2008).

This proposed study will attempt to address this gap by evaluating performance on standalone and embedded effort measures in people undergoing dementia assessments. McGuire et al.'s (2019) systematic review suggested that the Word Memory Test (WMT; Green 2003), Medical Symptom Validity Test (MSVT; Green, 2004) and Non-Verbal Medical Symptom Validity test (NV-MSVT;

Green, 2008) appear to be the most appropriate effort tests for use in neuropsychological assessment for dementia. McGuire et al. (2019) also note that Green's tests can be lengthy to administer therefore it is important to explore other options.

McGuire et al.'s (2019) review suggested that further research should evaluate the use of the two novel embedded indices within the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) – the Performance Validity Index (PVI) and the Charleston Revised Index of Effort for the RBANS (CRIER) created by Paulson and colleagues. Paulson et al. (2015) evaluated these indices in a non-dementia population and although they showed good potential as measures, they have suggested that replication is needed before considering clinical use. It will also further research by looking at a dementia population.

The aims of the proposed study are: (1) to examine what proportion of patients fall above/below the cut-off points for sub-optimal effort on Word Memory Test and the RBANS PVI and CRIER measures of effort; (2) to examine whether there is a significant association between performance on the WMT and each of the two RBANS measures (3) to examine the sensitivity and specificity of the RBANS measures to effort, when sub-optimal effort is defined by meeting criteria for sub-optimal effort on the WMT test. With regard to the association between performance on the WMT and the RBANS measures of effort, it is hypothesised that there will be a significant association between classification on the WMT and both the RBANS PVI and CRIER measures.

Proposed method

Participants

Participants will be patients undergoing neuropsychological assessment for possible dementia within NHS Highland. The service primarily sees older adult patients (over 65), though occasionally younger patients are referred for symptoms suggestive of early manifestation of dementia. Participants may be of any age up to 89 (RBANS and WMT tests are normed up to this age). Participants must have capacity to consent to participate in research.

There will be minimal patient exclusion criteria since the study aims to use a sample of all patients referred for neuropsychological assessment over the time period. However, patients should:

1. Have no significant hearing or vision problems which would prevent their completion of the neuropsychological tests.
2. Have English as their first language (for interpretation of test performance) and be able to read and write.
3. Not have a learning disability

Demographic information will be gathered; age, sex and years of education.

All eligible patients being assessed within a 9-10 month period will be invited to participate. In previous timeframes of this length, around 110 patients have been assessed for dementia in the service.

Following a clinical interview with a clinical psychologist or trainee, patients will receive neuropsychological assessment including effort tests to determine diagnosis. Data from all patients who receive a diagnosis of dementia or not will be included in the study and will be analysed separately.

Currently, it is difficult to assess how many patients are deemed to not be exerting optimal or suboptimal effort in this service, due to a lack of effort tests being used.

Measures

WMT will be used as the Gold Standard in this study to compare against the RBANS indices. Green (2011) has found that this test can achieve 98.4% specificity when testing for effort in a dementia population.

The BPS and AACN recommend the use of one standalone measure and at least one embedded measure of effort. The tests that will be used in this study are:

Standalone test:

The Word Memory Test (WMT, Green, 2003); a word list learning task that involves learning a list of 20 word pairs which are presented twice. It contains multiple subtests, of which, the first two are designed to measure effort.

Embedded tests: RBANS indices (PVI and CRIER) (Paulson, Horner & Bachman, 2015).

The RBANS (Randolph et al., 1998) is used in the assessment of dementia. The PVI and the CRIER are two novel embedded RBANS indices of effort.

Failure of these measures will be determined as specified in the test manuals. Evidence of poor performance on measures can be used as an index of insufficient effort.

The Geriatric Depression Scale (GDS) will also be administered as this is required to calculate the CRIER score (Yesavage et al., 1982).

The Geriatric Anxiety Scale (GAS) will also be administered to assess anxiety (Segal et al., 2010).

Procedures

Informed clinical consent will be obtained from each participant for undertaking a dementia assessment in the first appointment and a participant information sheet will be provided at the end of this

appointment. The participant will be asked to consider participation until the following week when testing begins, and research consent will be sought in the second appointment.

The neuropsychological battery administered by a psychologist/trainee will include RBANS and the WMT among other tests deemed clinically necessary.

WMT will be administered first. A time delay is needed for WMT (wait 20 minutes between the first and the second part) therefore part of the RBANS or any other test being used will be used in this timeframe. Tests used will be dependent on individual clinician's rationale and preference. Caution will be taken to avoid a word learning test during the delay due to interference effects.

Identification of valid or invalid responding will be based on WMT scores (using standard cut off scores).

Analysis

Proportions of patients meeting criteria for sub-optimal effort on each of the effort tests/indexes will be calculated. Secondly, Chi square tests, looking at the association in classification of performance on each of the effort tests (of those passing/failing each test) will be used. Finally, specificity and sensitivity of the measures will be determined and the area under the curve (AUC) statistic will be calculated using Receiver Operating Characteristic (ROC) analysis. WMT will be used as the gold standard against which the RBANS indices are examined for sensitivity, specificity but also Positive Predictive Value, Negative Predictive Value, Likelihood Ratio. ROC curves can be used to compare the diagnostic performance of two or more diagnostic tests (Griner et al., 1981).

Sample size

The sample size will be based on the Chi Squared analysis and calculated using G Power. It is anticipated that there will be relatively strong association between classification based on the RBANS and WMT measures of effort. In a memory disorders clinic sample, Paulson et al. (2015) used the TOMM together with the PVI and CRIER measures and found strong association. Using the TOMM to define sub-optimal effort, Paulson et al. (2015) reported the sensitivity and specificity of the PVI to be 0.82 and 0.77 respectively, whilst the CRIER index had a sensitivity of 0.84 and specificity of 0.90. These results suggest strong association between performance on the various measures of effort. Although the present study will use the WMT rather than the TOMM, the WMT has shown to have as good, if not better, performance than the TOMM in detecting sub-optimal performance and shown to have high specificity in a sample of people with dementia (Green et al., 2011).

Using G Power, with the effect size (Cohen's w) set at 0.5, alpha at 0.05, power at 0.8, df=1, the minimum sample size required is 32. To maximise power the aim will be to recruit at least 40 participants and more if time permits. As it is hypothesised that there will be a significant association

between classification on the WMT and both the PVI and CRIER measures, the alpha level will be p=0.05 for analysis in relation to each test.

A lack of existing data regarding the prevalence of sub-optimal effort in the context of neuropsychological assessment for dementia means that it is not possible to determine a sample size based on ROC analysis. The ROC analysis will be considered exploratory and prevalence data on sub-optimal effort that emerges from the present study may be useful in relation to future studies.

Ethical issues

Research on effort testing presents ethical issues in terms of how much can be explained to participants about the nature of the study without potentially impacting on performance. In this study, no tests will be used that are not currently recommended for routine clinical use. Participants will be asked to consent to data being used from the tests being analysed for the research. It will be explained to participants that the study is concerned with examining the accuracy of the tests used as part of their neuropsychological evaluation. They will be told that many different factors can affect how people perform on tests of memory and that we aim to ensure our tests are accurately measuring performance. Therefore, it will be explained that in this study, results will be compared from two of the tests used as part of their routine assessment to see how scores on the tests compare. It will further be explained that this will help increase our understanding whether the tests that we commonly use provide an accurate assessment of performance.

The project will be submitted to NHS ethics committee and NHS Research and Development Department for approval. During recruitment, participants will be informed of their right to withdraw their participation at any time. Data collection, storage and analysis will occur while following the principles of the Data Protection Act (1998) and GDPR guidelines.

Practical applications

This study will help to evaluate the use of novel RBANS indices when compared against WMT in a dementia population and to further expand knowledge where there is a paucity of research. If the novel RBANS indices are consistent with the WMT results, and show good specificity and sensitivity, then it will be suggested that these could be used as embedded tests of effort as they are relatively quick and easy to administer, therefore potentially improving the likelihood that effort testing be incorporated routinely in clinic.

Financial Issues

Expenses for neuropsychological measures will be needed.The WMT will be ordered from Green's Publishing as it is not available in the healthboard, Prof Evans will cover the cost from the MSc Clinical Neuropsychology budget.

Timetable

Stage	Dates	Timeframe
Ethics approval	March 2020	1-2 months
Data Collection	March 2020-December 2021	9 months
Submission	Mid Feb 2021	2 months for final write up
Viva	April 2021	

References

American Academy of Clinical Neuropsychology (AACN) (2007). American Academy of Clinical Neuropsychology (AACN) Practice Guidelines for neuropsychological assessment and consultation. *The Clinical Neuropsychologist*. 21, 209–231.

Boone, K. B., Savodnik, I., Ghaffarian, S., Lee, A., Freeman, D., & Berman, N. (1995) Rey 15-Item memorization and dot counting scores in a 'stress' claim worker's compensation population: Relationship to personality (MCMI) scores. *Journal of Clinical Psychology*, 51(1), 457 – 463.

Bortnik, K., Horner, M. & Bachman, D (2013). Performance on Standard Indexes of Effort Among Patients With Dementia. *Applied neuropsychology*.20 (10), 233-242.

BPS – British Psychological Society (2009). Assessment of effort in clinical testing of cognitive functioning for adults. Leicester, British Psychological Society.

Dean, A., Victor, T., Boone, K., & Arnold, G. (2008) The Relationship of IQ to Effort Test Performance. *The Clinical neuropsychologist.* 22 (10), 705-22.

Faust, D., Hart, K., Guilmette, T. & Arkes, H. (1988) Neuropsychologists' Capacity to Detect Adolescent Malingerers. *Professional Psychology: Research and Practice.* 19, 508-515.

Gervais, R., Rohling, M., Green, P., & Ford, W. (2004) A comparison of WMT, CARB and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology*, 19, 475 – 487.

Green, P.R., Lees-Haley, L.M., & Allen III.(2002) The word memory test and the validity of neuropsychological test scores. Journal of Forensic Neuropsychology. 97-124

Green, P. (2003). Word Memory Test for Windows: User's manual and program. Edmonton: Green's Publishing.

Green, P. (2004). Green's Medical Symptom Validity Test (MSVT) for Microsoft windows user's manual. Edmonton, Canada: Green's Publishing.

Green, P. (2008). Test manual for the nonverbal Medical Symptom Validity Test. Edmonton: Green's Publishing.

Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the word memory test and medical symptom validity test in groups with severe verbal memory impairment. *Applied Neuropsychology*, 18, 86–94.

Greve , K.W., Bianchini , K.J., & Doane , B.M. (2006) Classification accuracy of the Test of Memory Malingering in traumatic brain injury: Results of a known-groups analysis. J*ournal of Clinical and Experimental Neuropsychology*. 28, 1176 – 1190.

Griner P.F., Mayewski R.J., Mushlin A.I., & Greenland P. (1981) Selection and interpretation of diagnostic tests and procedures. *Journals of Internal Medicine.* 94:555-600.

McCarter, R. J., Walton, N. H., Brooks, D. N., & Powell, G. E. (2009). Effort testing in contemporary UK neuropsychological practice. *The Clinical Neuropsychologist*, 23(6), 1050-1066.

McGuire, C., Crawford, S., & Evans, J. (2019) Effort testing in dementia assessment: a systematic review. *Archives of Clinical Neuropsychology*, 34(1), 114-131.

Miele, A.S., Gunner, J.H., Lynch, J.K. & McCaffrey, R.J. (2012). Are embedded validity indices equivalent to free-standing symptom validity tests? *Archives of Clinical Neuropsychology*. 27(1),10-22.

Mittenberg, W., Pattonm C., Canyock, E.M., & Condit, D.C. (2002) Base rates of malingering and symptom exagerrration. *Journal of Clinical and Experimental Neuropsychology*. 24, 1094-1102.

Paulson, D., Horner, M., & Bachman, D. (2015). A Comparison of Four Embedded Validity Indices for the RBANS in a Memory Disorders Clinic. *Archives of Clinical Neuropsychology*. 30(3), 207-216.

Randolph, C., Tierney, M. C., Mohr, E., & Chase, T. N. (1998). The repeatable battery for the assessment of neuropsychological status (RBANS): Preliminary clinical validity. *Journal of Clinical & Experimental Neuropsychology: Official Journal of the International Neuropsychological Society.* 20(3), 310-319.

Segal, D. L., June, A., Payne, M., Coolidge, F. L., & Yochim, B. (2010). Development and initial validation of a self-report assessment tool for anxiety among older adults: The Geriatric Anxiety Scale. *Journal of Anxiety Disorder*. 24, 709-714.

Strauss, E., Sherman, E.M.S., & Spreen, O. (2006). A compendium of neuropsychological tests: Administration, norms, and commentary. New York: Oxford University Press.

Teichner, G. & Wagner, M. (2004). The Test of Memory Malingering (TOMM): Normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists.* 19, 455-64.

Tombaugh, T.N. (1996). The test of memory malingering. Toronto, ON, Canada: MultiHealth Systems.

Welsh, A.J., Bender, H.A., Whitman, L.A., Vasserman, M., & Macallister, W.S. (2012). Clinical utility of reliable digit span in assessing effort in children and adolescents with epilepsy, *Archives of Clinical Neuropsychology.* 27*(*7), 735-741.

Yesavage J. A., Brink, T. L., Rose, O., Lum, V., Huang, M., & V. O. Leirer. (1982) "Development and Validation of a Geriatric Depression Screening Scale: A Preliminary Report. *Journal of Psychiatric Research.* 17 (1), 37-49.

Appendix 2.8 Research cost form for original MRP

APPENDIX 8.6 RESEARCH COSTS & EQUIPMENT

RESEARCH EQUIPMENT, CONSUMABLES AND EXPENSES

Trainee MAIRI ORMBLE

Year of Course 2 Intake Year 2018

Please refer to latest stationary costs list (available from student support team)

Item	Details and Amount Required	Cost or Specify if to Request to Borrow from Department
Stationary	PAPER FOR INFORMATION SHEETS ETC	Subtotal: BORROW FROM DEPARTMENT
Postage	N/A	Subtotal:
Photocopying and Laser Printing	PAPER FOR INFORMATION SHEETS ETC	Subtotal: BORROW FROM DEPARTMENT
Equipment and Software	N/A	Subtotal:
Measures	WORD MEMORY TEST + BRAINS TBC	Subtotal: TBC
Miscellaneous	N/A	Subtotal:
Total		

For any request over £200 please provide further justification for all items that contribute to a high total cost estimate. Please also provide justification if costing for an honorarium:

Trainee Signature ... Date 21/10/19

Supervisor's Signature Date 22/10/19

Appendix 2.9 Health and Safety form for original MRP

APPENDIX 8.5 HEALTH & SAFETY FORM

HEALTH AND SAFETY FOR RESEARCHERS

1. Title of Project	COMPARISON OF EMBEDDED EFFORT TESTING TOOLS + A STANDALONE EFFORT TEST IN A DEMENTIA POPULATION
2. Trainee	MAIRI CRAMBIE
3. University Supervisor	PROFESSOR JON EVANS
4. Other Supervisor(s)	DR. JIM LAW
5. Local Lead Clinician	DR. ANN GALLOWAY
6. Participants: (age, group or sub-group, pre- or post-treatment, etc)	OLDER ADULT, NEUROPSYCHOLOGICAL AX. FOR DEMENTIA
7. Procedures to be applied (eg, questionnaire, interview, etc)	INTERVIEW + NEUROPSYCHOLOGICAL ASSESSMENT
8. Setting (where will procedures be carried out?) i) General	GENERAL, ROUTINE CLINIC FOR ASSESSMENT
ii) Are home visits involved	Y (N)
8. Potential Risk Factors Identified see chart	As this is routine clinical assessment, NO additional risks identified
10. Actions to minimise risk (refer to 9)	Supervision will be used routinely.

Trainee signature Date: 21/10/2014

University supervisor signature Date: 22/10/14

Appendix 2.10 – Information sheet for participants for original MRP

I***NFORMATION SHEET FOR PARTICIPANTS***

Title of project: A study of the accuracy of tests used in neuropsychological evaluations of memory

Invitation to take part in this study

We would like to invite you to take part in a research study. This sheet provides you with information to help you decide if you would like to be involved in the study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. All relevant contact details are at the bottom of this information sheet.

Who is conducting the research?

The research project is being conducted by Mairi Crombie (Trainee Clinical Psychologist), Dr Jim Law (Clinical Psychologist; NHS Highland) and Professor Jonathan Evans (Institute of Health and Well-being at the University of Glasgow).

What is the purpose of the study?

This study is concerned with examining the accuracy of tests used to assess memory as part of a neuropsychological evaluation. Many different factors can affect how people perform on these tests. We aim to ensure our tests are accurately measuring performance, therefore, in this study results will be compared from two of the tests used as part of the routine assessment to see how scores on the tests compare. This will help increase our understanding of whether the tests that we commonly use provide an accurate assessment of performance.

Why have I been chosen?

You have been chosen because you have been referred for a neuropsychological assessment to look at the difficulties you have been experiencing with memory.

Do I have to take part?

No, it is up to you to decide whether or not to take part***.*** You will be given this information sheet to keep until your next routine appointment. If you decide to take part, you will be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. If you decide to withdraw from the study, this would not affect your assessment or care in any way.

What will happen to me if I take part?

You will attend the clinic at Drumossie Unit in New Craig's hospital to complete some cognitive assessments which should last between 60-90 minutes. This is part of your routine assessment that would take place regardless of your participation in this study.

It will involve tests of concentration, memory and other cognitive functions as well as questionnaires about psychological wellbeing. All of the tasks administered in this study are standard neuropsychological tests. One of the tasks we will give you is not currently used in this department but is a recommended test. This does not take a substantial length of extra time to complete.

What do I have to do if I decide to take part?

You just have to attend your next assessment appointment as usual. We will ask you then if you would like to take part in the study. If you are happy to take part we will include the data from your assessment in the study.

If you prefer not to take part you should still attend for your routine clinic appointment. The person conducting your assessment will evaluate the results of your tests as usual, but we wont include any data from your assessment in our study.

What are the possible disadvantages and risks of taking part?

There are no disadvantages to taking part. As the tasks administered in this study are standard neuropsychological tests, there should not be any adverse effects from completing the tests.

What are the possible benefits of taking part?

The information collected in the study will give us a better understanding of factors that can influence the assessment of memory as part of a neuropsychological evaluation. This research could potentially benefit by ensuring that a true representation of your abilities and other client's abilities on neuropsychological testing is made.

Will my taking part in this study be kept confidential?

All information collected about you during the research will be kept strictly confidential. You will be identified by an identity number, and any information about you will have your name and address removed so that you cannot be recognised from it. Scientific publications arising from the research will not identify any individual.

What will happen to the results of the research study?

The results of the research will be written up but participants will not be identifiable. When the project is completed, the findings will be submitted for publication in peer reviewed international journals.

Who is organising and funding the research?

The research is organised and funded by the University of Glasgow.

Who has reviewed the study?

The project has been reviewed by the University of Glasgow College of Medical Veterinary and Life Sciences and by the NHS Highlands Ethics Committee.

Contact for Further Information

You can contact Mairi Crombie or Dr Jim Law who will be arranging and carrying out the assessments on 01463 253697; mairi.crombie1@nhs.net or jim.law@nhs.net or Professor Jon Evans on 01412110694 or Jonathan.Evans@glasgow.ac.uk who is supervising the research.

If you prefer to talk to someone who is not directly involved in the study about participating in research within the NHS you can contact Professor Tom McMillan

Professor Tom McMillan
Director of Research and Professor of Clinical Neuropsychology
Room 213 Level 2
Mental Health and Wellbeing
Gartnaval Royal Hospital
Glasgow G12 0XH
Email: Thomas.McMillan@glasgow.ac.uk
Telephone: 0141 211 0354

If you are unhappy about any aspect of the study?

We do not anticipate that there will be any reason for you to be unhappy whilst participating in this research, however, if you are about any aspect of the study and wish to make a complaint, please contact the researcher (Mairi Crombie) in the first instance. The normal NHS complaint mechanisms are also available to you, by calling 01463 705997

Thank you for considering this request to take part in the study.

Appendix 2.11 consent form for original MRP

CONSENT FORM

Title of project: A study of the accuracy of tests used in neuropsychological evaluations of memory

Patient Identification Number for the study:

Name of Researcher(s): Mairi Crombie

Contact details:

Drumossie Unit
New Craigs Hospital
Leachkin Road
Inverness
IV3 8NP
Email: mairi.crombie1@nhs.net

Phone: 01463 253697

Please write your initial in each box if you agree with the statement

I confirm that I have read and understand the information sheet dated x for the above study

☐

I confirm that I have had the opportunity to ask questions and that my questions have been answered sufficiently.

☐

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

I understand that I will not be able to be identified from the information in this study.

I understand that my information will remain held securely and will only be accessible to the research team and representatives of the study sponsor, NHS Highland (for audit purposes).

I agree to take part in the above study.

Participant Name	Date	Signature
……	… / … / ……	…….

Researcher	Date	Signature
……	… / … / ……	…….

Thank you for participating in this study.