New directions for analysis of process data from the Rey-Osterrieth Complex Figure Test

Tanya Kerr

Bachelor of Science (Honours), Master of Psychology, Master of Clinical Neuropsychology
31259693

Department of Psychology
Faculty of Human Sciences
Macquarie University, Sydney, Australia

Thesis submitted in partial fulfilment of the requirements for the degree of Master of Philosophy
2017

STATEMENT OF CANDIDATE

I, Tanya Kerr, candidate for the degree of Master of Philosophy at Macquarie University, hereby certify that:

i. This thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any other university or institution.

ii. I certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged.

iii. I certify that all information sources and literature used are indicated in the thesis.

iv. The research presented in this thesis was approved by the Macquarie University Ethics Review Committee (Human Research) (Ethics Ref: 5201600701)

Tanya Kerr

ACKNOWLEDGEMENTS

I gratefully acknowledge the support of my Co-Investigator Dr Jamie Berry who generously allowed me to use his research data and who played an instrumental role in this research project. I am deeply indebted to my primary supervisor, Associate Professor Jennifer Batchelor for her guidance and support. I also wish to thank my Co-Supervisor, Dr Susanne Meares for her encouragement and support.

I would like to thank Alan Warren for his invaluable IT assistance. I am thankful for the statistical oversight provided by Dr Naomi Sweller at Macquarie University, and further to Kim Colyvas, from the University of Newcastle for assisting with statistical analyses.

This project would not have been possible without the support of my family and friends. I will forever be grateful for the unconditional love of my parents who inspire me and have always encouraged me to learn. I am also indebted to my work family for supporting me throughout this project.

Lastly to my beautiful children, Ally and Josh, you are my greatest achievements. You both inspire me in ways that words cannot convey.

ABSTRACT

Process variables underlying copy strategy for the Rey-Osterrieth Complex Figure Test (ROCF) were examined in the context of assessing the cognitive construct of planning. While a number of tests of planning have been developed such as the Tower tests, the ROCF potentially facilitates a more detailed understanding of compromised planning performance. In the current study, quantified process measures were derived to evaluate individual copy approaches to the ROCF and were investigated relative to eight existing qualitative scoring systems. The performances of forty-nine healthy subjects were compared with the performances of fifty-two subjects in a mixed neurological sample. Quantified process measures and scores derived from existing qualitative scoring approaches were examined in relation to a range of psychometric properties including their capacity to produce scores which are normally distributed and to demonstrate variability between subjects. Discriminant validity was also examined, specifically whether process measures thus derived exhibit the ability to discriminate between normal and clinical subjects. Results demonstrated that novel quantified process measures can meaningfully contribute to analysis of performance parameters underlying copy of the ROCF. The relationship between copy approach and recall performance on the ROCF was also examined. Quantified process measures and scores from the existing qualitative scoring systems under study were found to be moderately correlated with recall across both normal and clinical subjects. Study findings were examined in the context of suggested future research and the importance of promoting the use of qualitative scoring systems in the analysis of suboptimal performance on the ROCF.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

AD	Alzheimer's disease
ACoA	Anterior Communicating Artery
ADHD	Attention Deficit Hyperactivity Disorder
BQSS	Boston Qualitative Scoring System
CAVLT	California Auditory Verbal Learning Test
CVA	Cerebrovascular Accident
D-KEFS	Delis-Kaplan Executive Function System
DSS-ROCF	Developmental Scoring System for the Rey-Osterrieth Complex Figure
FTD	Frontotemporal Dementia
HIV	Human Immunodeficiency Virus
IQ	Intelligence Quotient
MMSE	Mini Mental Status Examination
OCD	Obsessive Compulsive Disorder
PTA	Post Traumatic Amnesia
RCFT	Rey Complex Figure Test and Recognition Trial
ROCF	Rey-Osterrieth Complex Figure Test
TOL	Tower of London
WAIS-R NI	Wechsler Adult Intelligence Scale – Revised as a Neuropsychological Instrument
WMS-III	Wechsler Memory Scale – Third Edition

CHAPTER 1

Introduction

The Rey-Osterrieth Complex Figure Test

The Rey-Osterrieth Complex Figure Test (ROCF) is a widely used neuropsychological assessment measure (Rabin et al, 2005). In his comprehensive analysis of the utility of the ROCF as a clinical and research tool, Knight (2003a) writes: "It is a testament to the creativity and acumen of Andre Rey and Paul Osterrieth that their complex figure, administration process, and scoring system continue to be regarded as valuable components of neuropsychological assessments after 60 years" (p.25). The ROCF has been embraced by clinicians and researchers because it generates a variety of reproductions, informing our understanding of a diverse range of cognitive functions including spatial processing, graphomotor function, visual memory, perceptual organisation and planning.

The ROCF was originally developed by Rey (1941) and expanded by Osterrieth (1944). Corwin and Bylsma (1993) provide a translation of Rey's (1941) and Osterrieth's (1944) original articles in which administration and scoring procedures are detailed. The administration protocol adopted by Rey and Osterrieth involved a Copy trial and a 3-minute Immediate Recall trial. Subjects were instructed to copy the figure as accurately as possible. Although the copy trial was timed, no time restrictions were imposed. Coloured pencils were used and switched by the examiner, who also documented colour sequence in an effort to capture the process underlying individual reproductions of the figure. Without forewarning, subjects were

subsequently asked to recall the figure, a measure of information encoded incidentally.

While Rey (1941) accorded equal significance to the process by which the figure was copied as well as the accuracy of the reproduction, the former was not quantified. General information was provided to guide clinicians in their determination of poor reproductions – qualitative analysis. In contrast, the recall trial was scored for accuracy, with individual elements of the figure identified and scoring criteria developed, allowing a maximum possible score of 47 to be derived. Rey stressed the importance of evaluating recall scores in light of copy reproductions, thus allowing for a better delineation of cognitive processes contributing to scores thus obtained.

Using this same protocol, Osterrieth (1944, cited in Corwin & Bylsma, 1993) collected normative data on Copy and Immediate Recall performances for 295 children and adults aged from 4 years to 60 years. Results from this standardisation sample identified seven Reproduction Types, each posited to capture developmental variations and the evolving nature of the manner in which the complex figure is constructed. These reproduction types were defined briefly by Osterrieth as follows: (I) Construction of the central rectangle which serves as the foundation for the remaining elements; (II) Construction of one of the exterior details attached to the central rectangle, rectangle then completed and other details added; (III) General shape or outline of the figure drawn first, followed by internal details; (IV) Juxtaposition of details one following another in the absence of drawing the central rectangle, whole figure generally recognisable; (V) Recognisable details present in the

context of a confused background of lines; (VI) Reduction of the figure to a familiar scheme; and (VII) An unrecognisable scrawl (Osterrieth 1944, cited in Knight 2003b).

Osterrieth (1944, cited in Corwin & Bylsma, 1993) recognised that quantitative analysis of both copy and recall trials was essential in evaluating performance on the ROCF. He identified the numerous line segments used in Rey's scoring system to be potentially problematic, opting for a simpler scoring approach determined by the use of "structural elements" (p. 12). Eighteen elements were defined and scored according to accuracy and placement on a two-point scale. Two points were assigned to elements that were accurately drawn and properly placed. One point was awarded to distorted or misplaced elements. If an element was judged to be both distorted and misplaced, a half-point was given. Missing or unrecognisable elements were assigned zero points. Thus, the highest possible score was 36. Osterrieth's scoring system has largely remained unchanged with the exception of Taylor's (1959) adaptation, which provided more detailed scoring requirements relating to placement of a number of the 18 scoring elements. The Rey-Osterrieth scoring system has been the template for the development and evolution of different scoring approaches and will be referred to subsequently as the standard scoring system.

In the time since the ROCF was originally developed, a wide range of administration protocols have been formulated. Some use only copy administration (Visser, 1973) while more typically, Immediate and Delayed recall trials of varying intervals are used. While most employ an incidental learning paradigm, intentional learning protocols have also been examined (Tombaugh, Faulkner & Hubley, 1992).

One of the difficulties in making comparisons across studies is that different administration procedures are used, with some researchers including both Immediate (three minute) and Delayed recall trials, while others concentrate specifically on Delayed Recall. Loring, Martin, Meador and Lee (1990) demonstrated that Immediate recall trials can produce a facilitatory effect on both quantitative and qualitative aspects of Delayed recall performance in normal, healthy subjects. The timing of the Delayed Recall trial also varies between studies. Berry and Carpenter (1992) examined the effects of varying delay periods on recall performance of the ROCF following an Immediate Recall trial in a healthy aged sample. Findings demonstrated that Delayed Recall did not differ significantly for delay periods ranging from 15 to 60 minutes (Berry & Carpenter, 1992).

Inter-rater Reliability

The stability of scores across raters using quantitative scoring systems for the ROCF has been documented across a range of normal and clinical groups. Berry, Allen and Schmitt (1991) administered the ROCF as part of a larger battery of tests in a normal aged sample. ROCF protocols were scored according to a modified version of the standard scoring system where each of the 18 details was rated initially for distortion and then for displacement. Inter-rater reliability was calculated for a subset of 87 protocols. Reliability coefficients were reported to be significant across all three administration trials, that is, Copy ($r = 0.80$), Immediate Recall ($r = 0.93$) and 30-minute Delayed Recall ($r = 0.96$). Similar findings were reported by Berry and Carpenter (1992) in an elderly sample where the effect of four different delay periods on recall of the ROCF was examined. Reliability coefficients were also reported to be strong, Copy ($r = 0.95$), Immediate Recall ($r = 0.98$) and Delayed Recall ($r = 0.99$).

Tupler, Welsh, Asare-Aboagye and Dawson (1995) examined the inter-rater reliability of the standard scoring system in a clinical sample of elderly subjects, the majority of whom were diagnosed with dementia. Analyses revealed intraclass correlation coefficients of 0.93 for the Copy trial at the time of initial administration, and 0.94 when subjects were re-tested at a 3-month interval. Values for Delayed Recall trials were similarly strong (Delayed Recall Initial Administration: 0.94, Delayed Recall Re-Test: 0.96). Carr and Lincoln (1988) examined inter-rater reliability for the standard scoring system of the ROCF in 23 stroke patients and 17 general medical patients. Overall, a significantly high correlation (r= 0.99) was reported between total copy scores.

Loring, Martin, Meador and Lee (1990) assessed the reliability of the ROCF in a sample of college students using the standard scoring system. Inter-rater reliability was not calculated for the Copy trial given the restricted range of scores (>95% of sample scored 36/36). Inter-rater reliability for the Delayed Recall trial was reported to be 0.98. Liberman, Stewart, Seines and Gordon (1994) reported inter-rater reliability coefficients of 0.88 for the Copy trial, 0.97 for Immediate Recall, and 0.96 for Delayed Recall in a sample of 486 male amateur boxers. Statistically significant differences in mean scores between two raters across all three administration trials were, however, reported and some elements of the ROCF were associated with a higher rate of scoring differences than others. This is one of the first studies examining reliability of individual scoring elements of the ROCF, perhaps accounting for the discrepant findings reported. In a study evaluating memory performance in patients with complex partial seizures of temporal lobe origin, Breier, and colleagues (1996) reported strong interrater reliability for three indices of spatial and figural

memory derived from the ROCF. Reliability coefficients ranged from 0.87 to 0.94 across these indices.

Test – Retest Reliability

Multiple evaluations are generally undertaken in clinical assessments as a measure of decline or improvement in cognitive functioning. In this context, knowledge of test score stability over time is essential. Test-retest reliability of the ROCF is reported to be quite variable. In a sample of healthy elderly participants, Berry, Allen and Schmitt (1991) reported poor reliability for the copy trial ($r = 0.18$) and moderate reliability for the two recall trials (Immediate Recall: $r = 0.47$, Delayed Recall: $r = 0.59$) at 1-year re-test intervals for a subset of 41 participants in their healthy aged sample. Moderate reliability coefficients for the ROCF were reported at 1-year and 2-year re-test intervals by Mitrushina and Satz (1991) in a group of healthy, aged participants. Values ranged from 0.56 to 0.68 for Copy trials and 0.57 to 0.62 for Delayed recall. Meyers and Meyers (1995) provided test-retest data for a subset of their normative sample where sufficient range in scores was evident. Evaluation of these 12 subjects revealed reasonable reliability for both Immediate Recall ($r = 0.76$) and Delayed Recall ($r = 0.89$). Tupler, Welsh, Asare-Aboagye and Dawson (1995) reported intraclass correlations of 0.94 for Copy and 0.95 for Delayed Recall at a three-month re-test interval in an elderly sample of subjects, the majority of whom had been diagnosed with dementia.

An examination of the literature on the temporal stability of the ROCF, identifies a number of issues (Knight, 2003b). A confounding factor is the loss of novelty following initial exposure of the figure and the shift from an incidental recall

paradigm to an intentional recall paradigm in cases where the examinee recalls the stimulus or aspects of the administration process (Tombaugh & Hubley, 1991; Tombaugh et al, 1992). There is further potential for examinees to benefit from initial exposure to the test and improve their approach on re-test. In their investigation of practice effects in repeated administration of a range of neuropsychological tests, Mitrushina and Satz (1991) reported that serial visual memory assessment adopting both Immediate and Delayed Recall trials, yielded observable practice effects across four age groups in their subject sample spanning 57 to 85 years, with a lower magnitude of practice effects evident as age increased. The subjects in this study were well educated. Levine et al., (2004) reported modest effect sizes for recall trials of the ROCF in 478 healthy, largely Caucasian males (Mean Age: 42.2, SD=8.6) who were well educated (Mean Education: 16.4, SD=2.3). The retest interval ranged from 4 to 24 months. Practice effects were not demonstrated on copy trial, though the authors suggested there may have been a ceiling effect on copy of the ROCF in this sample. A regression equation was developed for estimation of copy and recall scores on retest. The length of the retest interval and educational level were not found to contribute significantly to the regression equation (Levine et al, 2004). More research is needed to examine the stability of ROCF scores over time. This is especially important in clinical settings where changes in ROCF scores are considered alongside changes across other cognitive measures in informing decisions regarding diagnosis and management.

The availability of equivalent, alternative complex figures is important to address concerns relating to the confounding influence of exposure of the same stimulus in review assessments. The Taylor Complex Figure was initially developed

by Taylor (1969) as a postoperative memory test in research on the effects on nonverbal memory of right and left temporal lobectomies. It was assumed that the Taylor Complex Figure was comparable to the ROCF, which was used as the pre-operative measure (Taylor 1969). The assumption of comparability of the two figures is based on the fact that both have an equal number of elements. This assumption, however, has not been supported across a wide range of clinical and non-clinical groups. Strauss and Spreen (1990) administered Copy and 30-minute Delayed Recall trials of both figures in a sample of college students. While performances on the copy trials were comparable using the standard quantitative scoring system, a 5-point difference was observed in favour of recall scores on the Taylor figure. This result was replicated by Tombaugh and Hubley (1991) who also examined comparability of these figures in a sample of college students. Across a range of variables encompassing length of delay interval, incidental versus learning paradigm, scoring system used and size of figure, it was consistently revealed that while parallel findings were demonstrated on the copy trial, the Taylor figure was easier to recall than the ROCF. Other studies have further demonstrated that while the figures yield equivalent copy scores, the Taylor figure is more susceptible to verbal mediation (Casey et al, 1991) and is easier to recall (Tombaugh et al, 1992; Hamby et al 1993). Delaney, Prevey, Cramer, and Mattson (1992) examined the comparability of both figures in a normal sample. Again, while the performance of subjects on copy administration for both figures was comparable, significantly better performances were evident on the Taylor figure for both immediate and 20-minute delayed recall. These findings have been replicated by Duley, Wilkins, Hamby, Hopkins, Burwell and Barry (1993) in a clinic sample where Copy and 30-minute Delayed Recall trials

were administered to a group of HIV patients. Superior recall for the Taylor Figure was observed at both Immediate and Delayed recall trials.

Demographic Variables

Differences in the demographic characteristics of participants across studies of the ROCF makes it difficult at times to ascertain which factors are responsible for differences in scores observed. The significant relationship between age and performance on the ROCF has been well documented, with performance deteriorating as age increases. Age effects are more consistently reported for recall trial performance than for copy of the ROCF in both normal samples (Bennett-Levy 1984; Berry et al 1991; Boone et al, 1993; Chiulli et al, 1995; Ostrosky-Solis et al, 1998; Fastenau et al, 1999; Rosselli & Ardila, 1991) and clinical groups (Powell, 1979; King, 1981). In a non-clinical sample aged 16 to 69 years, Gallagher and Burke (2007) found that performance decline on the ROCF emerges in the "late forties/early fifties" (p. 42). Hartman and Potter's (1998) qualitative analysis of performance differences between healthy younger subjects (age range: 18-32 years) and healthy elderly subjects (age range: 60-81 years), revealed a larger frequency of minor errors in the drawings of older subjects on the copy trial. Age related differences in recall were also documented, largely involving loss of information rather than distortion of figure elements. Similar findings were observed by Janowsky and Thomas-Thrapp (1993) in a normal elderly sample who performed significantly worse on recall of the ROCF than younger subjects, despite comparable ability relative to younger subjects on the copy trial. These findings have been replicated in a non-clinical elderly sample by Mitrushina, Satz and Chervinski (1990).

Gender differences in performance on the ROCF have been less consistently reported. Gender was not found to be a unique predictor of performance on qualitative measures of the ROCF derived by Bennett-Levy (1984) in a healthy sample, despite the fact that gender was significantly associated with copy and recall performance. Gender also contributed minimally to performance on the ROCF in studies using normal subject samples (Boone et al,1993; Chiulli et al 1995; Janowski & Thomas-Thrapp 1993) and for clinical groups (King 1981). It is generally accepted that while gender differences may be observed in research to date, the unique variance contributed by gender to performance on the ROCF is negligible (Berry et al, 1991; Chiulli et al, 1995; Fastenau et al 1999). This is disputed by Gallagher and Burke (2007) who reported significant effects for gender in a normal sample aged 16-69 years, with males consistently outperforming females. Gender differences in performance on the ROCF were also reported by Rosselli and Ardila (1991) in a non-clinical Spanish speaking sample.

As with gender differences, mixed findings have been reported regarding the influence of education on performance of the ROCF. Bennett-Levy (1984) reported that scores on the ROCF were significantly related to estimated IQ in a non-clinical sample aged 17-49 years. Research findings in support of the effects of education on ROCF scores in healthy subjects have also been reported by Berry et al (1991) and for a modified version of the ROCF, the Extended Complex Figure Test, in which recognition and matching trials were included in the administration (Fastenau et al, 1999). Ponton and colleagues (1996) administered the ROCF as part of a larger test battery to Spanish speaking healthy subjects and scored Copy and 10-minute Recall reproductions according to the standard scoring criteria. A floor effect

was observed in less educated groups and it was argued that the ROCF may represent an inadequate measure of visuospatial construction in poorly educated individuals. Similar findings were reported by Ardila, Rosselli and Rosas (1989) who compared performances on visuospatial tasks including the ROCF across well educated and illiterate Spanish speaking subjects matched for gender and age. In their standardisation sample for the Rey Complex Figure Test and Recognition Trial (RCFT), Meyers and Meyers (1995) reported that education only accounted for 2% of the variance on average of age-corrected ROCF scores, leading to their decision to omit an education correction in their normative data for this standardised scoring system. Their normative sample comprised 601 adults aged 18-89 years, from which a subset of 394 individuals were selected to create a demographically corrected sample. While a linear effect for age was observed across identified RCFT variables, there was no significant relationship between years of education and age-adjusted scores on any of these variables. The average education level of the subsample (M=13.91, SD=2.48) was noted to be a little higher than the US population at the time (Meyers & Meyers, 1995).

Ashton et al (2005) also reported that education was not significantly predictive of the variance in copy and recall performance on Meyers and Meyer's (1995) RCFT in 100 patients with traumatic brain injury. Education levels of subjects in this study ranged from 8-18 years, (Mean = 12.4). Delaney et al's, (1992) study of test-retest reliability for the ROCF and Taylor Complex Figure in a non-clinical sample with education levels ranging from 6-16 years (Mean = 12.8), did not find significant correlations between education and performance on the ROCF. Correlations with performance across copy and recall trials were reported to range

from -0.01 to -0.2 (Delaney et al, 1992). Similar findings were reported by Boone et al., (1993) in a well educated, elderly normal sample where gender and education were not unique predictors of performance on Copy, Immediate Recall and Percent Retention over 3-minute delay.

Overall, there is strong support for the influence of age on performance of the ROCF and less consistent evidence for the effects of gender and education. This is reflected in published scoring systems for the ROCF where normative data is corrected for age, but not for other demographic variables (Meyers & Meyers, 1995).

Qualitative Scoring Systems for the Rey Osterrieth Complex Figure Test: An Analysis of Existing Approaches

The term *qualitative* is an amorphous construct which has been interpreted variously when applied to neuropsychological assessment. Within the context of evaluations of performance on the ROCF, the term *qualitative* is primarily used to refer to processes variables underlying copy strategy for the complex visual stimulus. Existing scoring systems attempting to capture copy approach for the ROCF have been widely referred to as qualitative scoring systems.

It is not altogether surprising that Rey (1941) and Osterrieth (1944) were the first to recognise the importance of qualitative variables in their process analysis of reproductions of the complex figure. In fact, both Rey and Osterrieth highlighted the importance of the approach taken to copying the figure as well as the final copy. Osterrieth proposed that organisation of the copy was strongly related to

recall performance. He identified ‘Reproduction Types’ and presented normative data for both copy and recall trials. While the standard quantitative scoring system thus developed emphasised the accuracy of the appearance and placement of elements on the ROCF, it has come under criticism for failing to precisely characterise qualitative aspects of performance. Over the years, research has focused on qualitative scoring systems which generate variables reflecting the approach and strategy used in reproducing the ROCF. It is argued that analysis of the process by which test responses are generated identifies important information regarding the numerous underlying pathways contributing to observed scores (Kaplan, 1988). Several scoring systems have been developed to characterise approaches to copying the ROCF. Many of the qualitative scoring systems have been developed to highlight the importance of planning and organisation strategies on copy and recall of the ROCF, and assist in discriminating between poor planning/organisation and poor recall.

A number of scoring systems have been developed to capture process variables across both paediatric and adult populations. Paediatric populations will not be examined in detail given the focus on adult subjects in the current study. The developmental literature does, however, inform our understanding of age related changes in organisational strategies and the continuing development of these strategies from early childhood through to at least middle adolescence (Anderson et al, 2001; Akshoomoff & Stiles, 1995; Waber & Holmes, 1985; Waber & Holmes, 1986). As is observed with adults, children who adopt an organised approach while copying the ROCF are more likely to demonstrate stronger recall in comparison to children who use a fragmented approach (Anderson et al 2000).

One of the only comprehensive investigations of existing qualitative scoring systems was undertaken by Troyer and Wishart in 1997. They compared the psychometric properties of nine qualitative scoring systems relevant to their capacity to evaluate strategy and organisation for copy and recall trials of the ROCF in a healthy sample of high functioning older adults aged 60 to 91 years. Scores derived for each of these systems were examined across parameters including distribution of scores, relation to memory performance, construct validity, and inter-rater reliability. These psychometric characteristics were deemed to be essential for the identification of individual differences and changes in performance over subsequent assessments, qualities which are of integral importance in the clinical context. All scoring systems were found to demonstrate moderate to high discriminant validity and inter-rater reliability. Differences were evident, however, in relation to distribution of scores, recall, and convergent validity. Seven of these scoring systems will now be discussed in more detail as they will form the basis of comparison for the quantified process measures developed in the current study. Two of these scoring systems, that of Hamby, Wilkins & Barry (1993) and the two indices from Stern et al's (1999) Boston Qualitative Scoring System (BQSS) will not be addressed given the significance those systems place on accuracy of copy reproductions, which was not the target of analysis in the current study. Additionally, a qualitative scoring system developed by Savage and colleagues (1999) will also be examined.

Visser (1973)

Visser (1973, cited in Knight 2003b) developed a qualitative scoring system for the ROCF, primarily in the context of assessing the impact of acquired brain impairment on reproductions of the complex figure. The sequence of element

construction was identified as the index of impairment. Visser's Complex Figure Test includes a copy trial only, and the figure stimulus is presented in a 90-degree counterclockwise orientation. There is no reason provided as to why the portrait orientation was favoured. Drawings are scored based upon the individual order in which each of 35 identified lines from the figure are drawn. Each line is given a sequence number based upon flowchart representations of how the ROCF is copied.

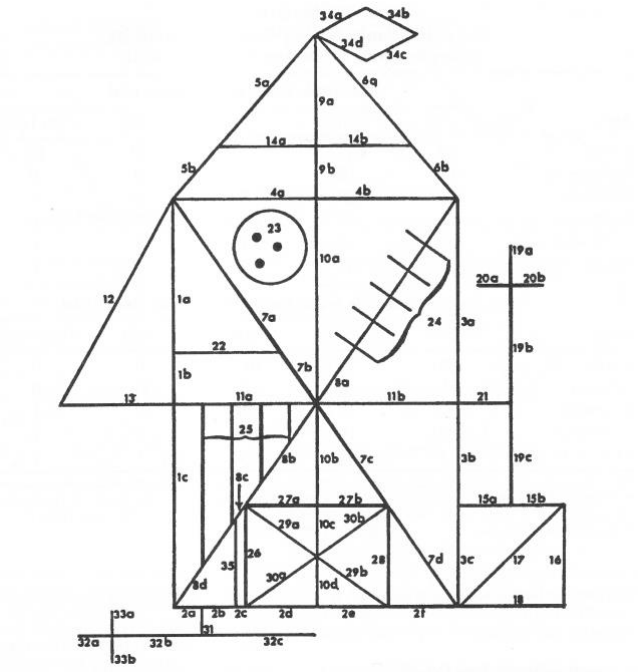

Figure 1. Sequence lines forming the Visser Complex Figure [Source: Knight (2003b), p. 136]

Three categories of scores are derived: Interruptions (drawing a line in two or more parts, separated by at least one other line), awarded 1 point; Omissions (lines are absent from the drawing), awarded 1 point; and Sequence Items (1 point awarded whenever one of 17 sequence conditions are met). A Total Score is calculated by combining the three category scores. High scores denote poor performance and are suggested to provide a measure of fragmentation.

Visser's Complex Figure Test was standardised on a heterogeneous group of 328 neurologic patients and 247 normal controls (140 males, 107 females). A 63.4% classification rate for discriminating acquired brain impairment was reported. Copy trial scores were reported to be correlated with age (r=0.22). Test retest reliability using intervals of 1-2 weeks was reported to be 0.84 in a clinical sample of 17 males with alcohol dependence (Visser, 1973, cited in Knight 2003b). Significant correlations between scores based on Visser's scoring system and scores on the Embedded Figures Test (r=0.48) and the Block Design subtest from the WAIS (r=0.34) were reported in the clinical sample and normal controls, providing supportive evidence for convergent validity (Visser, 1973, cited in Knight, 2003b).

One of the advantages of Visser's qualitative scoring system is that it systematically examines the sequence in which lines are drawn as a unique process variable and this has also been adopted in subsequently developed scoring systems under examination. A significant limitation is represented by the novel administration, thus limiting comparability with findings from protocols adopting the standard administration of the ROCF, though figure orientation has been suggested to have little influence on ROCF performance (Ferraro, et al, 2002). Scores obtained using this qualitative scoring system were found to demonstrate negative skewing in Troyer and Wishart's (1997) comparison of qualitative scoring systems, reflecting clustering of scores at high values.

Binder (1982)

Binder (1982) developed a qualitative scoring system identifying five elementary units in the ROCF –the horizontal midline, the vertical midline, the two diagonals, and the vertex of the triangle. Like Visser (1973), Binder emphasised the importance of documenting the sequencing strategy adopted when copying the ROCF. A 'Configural Score' (Range: 0-5) is calculated according to how the five elements of the figure are constructed. Scoring rules are dichotomous and credit is only given if elements are drawn as a continuous line segment. Credit is not granted if units are drawn in a fragmented manner. The Configural Score is intended to represent a qualitative index of fragmentation. Binder reported that there was 100% agreement between two independent raters regarding classification of drawings as fragmented using this scoring approach (Binder 1982).

Binder (1982) used this qualitative scoring system to examine copying strategies adopted by 28 patients with unilateral stroke subdivided into equal groups of right brain lesions and left brain lesions, compared with 14 healthy controls matched for age and education. Patients with unilateral stroke performed more poorly, both quantitatively and qualitatively when compared with control subjects. Control subjects made accurate reproductions and drew segments comprising the Configural Units without evidence of fragmentation. Although patients with right hemisphere lesions performed most poorly, with increased evidence of distortions and left-sided neglect, patients with left hemisphere lesions also demonstrated a fragmented, piecemeal approach, though the final copy was reasonably accurate (Binder 1982).

Binder's scoring approach has attracted criticism since it fails to incorporate much of the information available in the ROCF, given that most elements are not rated in this system (Hamby, Wilkins & Barry, 1993; Shorr, Delis & Massman, 1992). Another criticism is that the 0-5 range of the Configural Score may not be sensitive to variations in performance across a range of clinical groups (Shorr et al, 1992) The all or none fashion of scoring designated units has also been questioned with the suggestion that it might lead to scores which under-represent an individual's true performance (Shorr et al, 1992). Troyer and Wishart's (1997) psychometric analysis of qualitative scoring systems further identified that Binder's scoring system produced significantly skewed scores. Not withstanding these criticisms, Binder's scoring approach has formed the mainstay of a number of qualitative scoring methods which followed.

Bennett-Levy (1984)

Bennett-Levy (1984) developed scoring measures intended to quantify copying strategy based on Rey and Osterrieth's standard scoring system. This scoring approach is defined by Gestalt principles of perceptual organisation (Wertheimer, 1958), characterised by the order in which elements are sequenced during construction of the drawing and the degree of fragmentation of elements. This parallels Visser's scoring approach. Symmetry and good continuation, two of the principles of perceptual organisation as outlined by Wertheimer (1958), are identified as central factors in deriving strategy scores. Good continuation is defined as a straight line drawn continuously in one segment until it reaches correct intersection with another line. Seventeen points of good continuation are outlined, as well as one point of poor continuation. The maximum Continuation Score is 18.

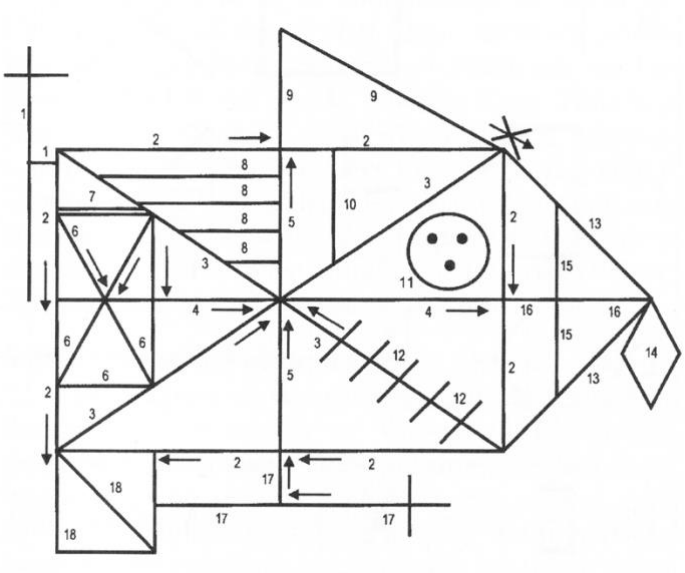

Figure 2. Line sequences defining Bennett-Levy's Continuation Score [Source: Bennett-Levy (1984), p. 112]

The principle of symmetry is assumed to reflect the structure and symmetry perceived within the figure as defined by the order in which component features are drawn. Points are given for successive construction of symmetrical units and their symmetrical components. The maximum symmetry score attainable is 18 points. A Strategy Total score is calculated by combining the Good Continuation and Symmetry Scores. While a theoretical maximum of 36 points exists, this can never be attained given that good continuation and symmetry strategies are in direct conflict at three points during construction of the drawing.

The 'Strict' scoring approach adopted by Taylor (1959) was used by Bennett-Levy to rate Copy trial drawings, where accuracy is defined in terms of presence, distortion and misplacement of figure elements. Draftsmanship is also considered. Copy strategy scores are intended to complement Copy accuracy ratings. Recall performance is scored according to 'Strict' and 'Lax' scoring procedures. In

the latter system, distortions and misplacements are scored less strictly and 'tidiness' of the reproduction is noted to be unimportant (Bennett-Levy, 1984).

Bennett-Levy (1984) administered Copy and 40-minute Delayed Recall trials of the ROCF to 107 healthy subjects (age range: 17-49 years). Inter-rater reliability for 25 randomly selected protocols was reported to be 0.96. The Lax Recall score as calculated by two independent raters for these 25 randomly selected recall drawings yielded an inter-rater reliability coefficient of 0.98. Inter-rater reliability scores are, however, not provided for Bennett-Levy's measures of Symmetry and Good Continuation.

Bennett-Levy (1984) reported that copy strategy measures derived from this qualitative scoring system, both individually and in combination, predicted copy accuracy and recall performance on the ROCF in a normal population. Copy Strategy scores were also reported to differentiate primary memory deficits from scores reflecting poor organisation during reproduction of the figure. While estimated intellectual ability was significantly correlated with copy and recall of the ROCF, multiple regression analyses revealed that the effects of copy strategy were independent of estimated intellectual ability.

One of the difficulties with Bennett-Levy's scoring system is its complexity. Bennett-Levy's Copying Strategy has also not been validated in clinical populations. This scoring system featured strongly, however, in Troyer and Wishart's (1997) psychometric analysis of available qualitative scoring systems. It was demonstrated to produce a wide range of scores, and further, scores were found to be

normally distributed. Bennett-Levy's scoring system was also shown to produce scores that were moderately related to recall scores (Troyer & Wishart, 1997). As with Visser's scoring method, this approach incorporates sequencing of individual elements of the ROCF as a central measure of planning.

Shorr et al (1992)

Shorr, Delis and Massman (1992) devised a qualitative scoring system for the ROCF intended to quantify "the use of an organised strategy for copying numerous ROCF subwholes and isomorphic features from the same perceptual category" (p.46). Shorr and colleagues elaborated on Binder's (1982) qualitative scoring system, opting for a continuous scoring system in place of a dichotomous scoring approach. This scoring system incorporates a measure of perceptual clustering, a construct used by Delis (1989) as the visuospatial equivalent of semantic clustering. Perceptual clustering involves organisation of the numerous elements of the ROCF into a smaller, more manageable number of perceptual units, thus enabling more efficient encoding and retrieval. Four separate scores are provided for reproduction and recall of the ROCF. A score for overall accuracy of the ROCF is computed using the standard scoring system. A Perceptual Cluster Score is also calculated for each copied drawing. Flow charts of figure copies are used to compute a Perceptual Cluster Index, comprised of subwholes of the ROCF. These include the central rectangle and its substructures - diagonals, horizontal and vertical midlines; the vertices of the triangle attached to the central rectangle; the small rectangle within the central rectangle and the diagonals within this structure; and the small square attached to the bottom of the central rectangle. For each subwhole, junctures are identified where breaks in continuous drawing can occur, leading to the development

of a scale assessing how well individuals organise the figure. In this system, 20 junctures in 8 organisation units are identified.

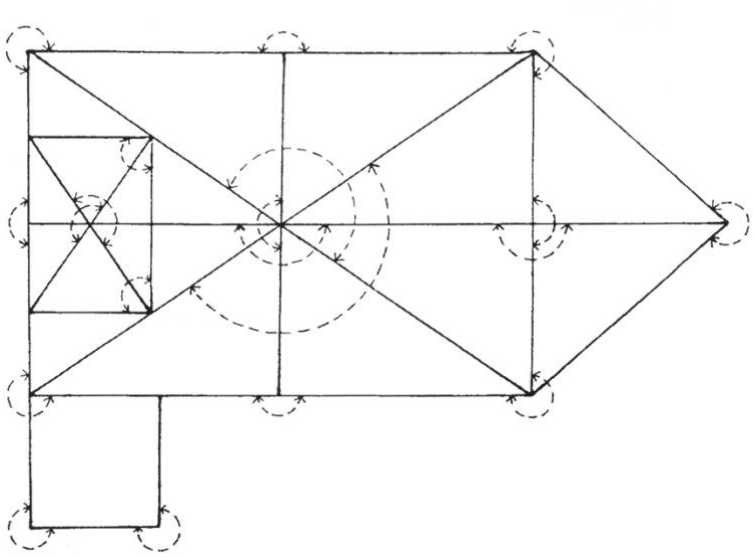

Figure 3. Identified junctures for Perceptual Cluster Index [Source: Shorr et al, 1992]

Credit towards the perceptual cluster is given if a line on each side of the juncture is drawn continuously or contiguously, thus enabling measurement of "subtle variations in perceptual clustering" (p.46). The maximum Perceptual Cluster Score is 20 points. A Perceptual Cluster Ratio, which controls for the impact of missing junctures, is computed by dividing the Perceptual Cluster Score by the total number of junctures present. An Encoding Score is additionally provided to capture level of encoding and to control for construction ability. This is calculated by dividing the Immediate Recall Accuracy score by the Copy Accuracy score. A Savings score (Delayed Recall Accuracy divided by Immediate Recall Accuracy) is also computed and denotes the percentage of information from Immediate Recall, remembered at Delayed Recall.

This qualitative scoring system was used in a sample of 50 neuropsychiatric patients, where inter-rater reliability for 15 randomly selected patient drawings was reported to exceed 0.98. Memory performance across this patient group was found to be more strongly correlated with the Perceptual Cluster Ratio rather than the standard copy accuracy score. This trend was noted to be independent of the degree of visuospatial deficits across the subject sample. The Perceptual Cluster Ratio was also found to contribute uniquely to the Encoding Score (Shorr et al, 1992). Troyer and Wishart's (1997) psychometric analysis of this qualitative scoring system found derived scores to be negatively skewed, identifying clustering of average scores in the maximum range. This draws scrutiny on the inclusion of additional elements from the ROCF, which the authors argued would better define scoring approach. As with Bennett-Levy's scoring method, emphasis is placed on the continuation of individual elements through to completion, identifying fragmentation as a feature of suboptimal performance.

Bylsma et al (1995)

Bylsma, Bobholz, Schretlen and Correa (1995) reported on a scoring system for the ROCF which provides a "quantitative" index of how individuals copy the figure - the Q-Score. This qualitative scoring system also relies heavily on the premise that drawing the structural elements first, followed by addition of the details, represents the most efficient copying approach reflecting good planning. The scoring index is based on the order of production of structural elements using contiguous lines. The examiner is required to document each line of the subject's copy sequentially in order for the Q-Score to be derived. Bylsma (2008) elaborates upon this scoring system in more detail. Thirteen units, each comprising two or more lines

are identified. The Central Rectangle, Vertical and Horizontal Bisectors, and the Diagonals within the Central Rectangle are accorded more weight given their assumed significance in producing an accurate copy. Order points are additionally credited if the Central Rectangle is completed first and the Bisectors and Diagonals represent the second and third units completed. Unit and Order scores are combined to provide the Q-Score, an index of the individual's planning and strategic approach to copying the ROCF.

Bylsma and his colleagues (1995) report that this scoring system has excellent inter-rater reliability (r=0.99), and that Q-scores demonstrate greater sensitivity relative to accuracy scores in discriminating evaluations of ROCF copy strategy by independent clinicians. Factor analysis of the scale was reported to identify 5 factors (Bylsma et al, 1995). Factor scores were reported to be correlated with measures of executive function such as the Trail Making Test, Part B ($r = -0.14$, $p<0.02$) and phonemic fluency ($r = 0.23$, $p<0.03$) (Bylsma et al, 1997). Troyer and Wishart's (1997) analysis revealed Bylsma's system to demonstrate normally distributed scores and further, a wide range of possible scores. The Q-Score was also found to be moderately related to recall scores. As with Binder's (1982) and Shorr et al's (1992) scoring methods, this scoring system foregrounds the configural elements of the ROCF in copy strategy.

Savage et al (1999)

Savage, Baer, Keuthen, Brown, Rauch and Jenike (1999) developed a qualitative scoring system for the ROCF based on the configural elements defined by Binder (1982), with the addition of a fifth unit, the central rectangle, each of which

has to be drawn in an unfragmented manner in order to receive credit for organisation. Organisational Strategy is defined both by quantitative and descriptive analysis of organisational sequence. The order of construction and drawing accuracy is not considered in the quantitative analysis, which involves assignment of 2- points to the central rectangle, identifying its importance to the organisation of the figure and 1-point for each of the four designated elements. Scores therefore range from 0-6, with higher scores indicating stronger organisational ability. Descriptive analyses of organisation sequence identify approaches which involve initial and subsequent construction of details (components other than the 5 basic configural units), thereby providing information pertaining to "early organisational sequence" (p. 908). Construction Accuracy is determined using a scoring system developed by Denman (1984) where 24 segments of the figure are evaluated on the basis of information relating to sector location, line angles, line length, and line number. This index was not examined in the current study where the emphasis was on measures of planning. Inter-rater reliability for the Organisational Strategy score is reported to be high (r=0.96) (Savage et al, 2000).

Savage and colleagues (1999) examined the influence of organisation strategy on visual memory in 20 unmedicated patients with Obsessive Compulsive Disorder (OCD) and 20 normal controls matched for age and education. OCD patients demonstrated significant impairments on measures of strategic organisation and immediate non-verbal memory relative to controls. It was proposed that identified impairments reflected impaired encoding of information during copy, without evidence of rapid forgetting. These findings were replicated in an independent sample of OCD patients, with findings further extended to a verbal learning task, adding

support to the conclusion that patients with OCD demonstrate difficulty on memory tests placing demands upon organisational and strategic processing (Savage, Deckersbach, Wilhelm, Rauch, Baer, Reid and Jenike, 2000).

This qualitative scoring system was also used to characaterise organisational approach in reproductions of the ROCF in 71 patients diagnosed with Obsessive-Compulsive Disorder (OCD) and 55 age matched healthy controls (Deckersbach et al, 2000). When the psychometric properties of this system were compared with Shorr et al's (1992) scoring system, inter-rater reliabilities were comparable for both systems, and organisation scores during copy were noted to account for a significant portion of the variance in free recall scores across both approaches. It was reported that not all organisation elements of Shorr et al's (1992) scoring system were equally predictive of memory performance, suggesting the merits of Savage et al's (1999) simpler scoring system. The authors did, however, highlight the narrow focus of their study on whether organisational features were drawn as whole units, and suggested that other aspects of organisation might equally predict recall performance. They also reported that regression analysis did not support the weighting of the central rectangle in adding significance to the prediction of accurate delayed recall, thereby suggesting that it was accorded 1-point instead of 2 points, resulting in a 5-point scoring system.

The psychometric properties of the refined Savage-Deckersbach scoring system (5-point scoring system) have been examined in a sample involving undergraduate students and participants drawn from a university-based neuropsychology assessment clinic (Smith et al, 2007). This system was reported to

demonstrate adequate internal consistency, with superior inter-rater reliability when compared to the standard scoring system. Scores derived from this qualitative scoring system did not, however, differentiate the two groups, which the authors attributed to the composition of the clinical group, largely comprising high functioning individuals with difficulties of learning and attention.

Comparison of this scoring system with Binder's (1982) method will provide important information regarding the significance of including the central rectangle. The central rectangle formed an important part of Osterrieth's original qualitative framework (Osterrieth, 1944) and also comprises the configural framework for scoring procedures developed by Shorr et al (1992) and Bylsma (1995).

Waber & Holmes (1985)

Waber and Holmes (1985, 1986) developed a qualitative scoring system for the ROCF for use in paediatric populations, where obtained scores are considered in the context of developmental changes. This has since been published as the Developmental Scoring System for the Rey Osterrieth Complex Figure – DSS-ROCF (Bernstein & Waber, 1996). In the original normative studies (Waber & Holmes, 1985, 1986), the ROCF was administered to 454 normal children aged 5-14 years. Copy, Immediate and Delayed Recall trials were administered. Coloured pens were used to capture the sequence of the drawing process. Four qualitative aspects of ROCF performance were assessed using this scoring approach: Organisation, Style, Accuracy and Errors. A gradual increase in organisational scores and configurational strategies between the ages of 5 and 14 years was observed using this scoring

approach. The authors reported that children developed the ability to reproduce all features of the ROCF by the age of 9 years and that beyond that age, performance changes reflected varying levels of efficiency in planning and organising construction of the ROCF (Waber & Holmes, 1985). In relation to recall performance, salient findings were noted to include that across all age groups, organising structures of the ROCF were better recalled than figure details, and errors and distortions were more evident on recall than on copy, regardless of memory delay. With the exception of the youngest subjects, drawings on recall were noted to demonstrate a more configurational style relative to copy trials (Waber & Holmes, 1986).

In the DSS, the ROCF is broken down into four structural units in order for accuracy to be scored: a Central Rectangle comprising 12 segments, the Main Substructures comprising 13 segments, the Outer Configuration structures comprising 26 segments, and the Internal Details comprising 13 segments. Line segments are coded as either present or absent. Accuracy Scores are also provided, representing the sum of line segments present. Rather than a total accuracy score, accuracy subscores are generated, enabling comparison of the presence of configural versus detail elements of the figure. Twenty four criterial features, assumed to define "goodness of organisation" (p. 563), are scored as present or absent for the copy trial, and 16 for each of the recall trials. These criteria are intended to reflect either alignments or intersections of designated line segments. This information is used to assign drawings to one of five levels of organisation and subsequently, categorisation on the basis of style within each organisational level. Four types of error patterns are recorded for copy and memory productions: conflation, rotation, perseveration, and

misplacement. An Error Score is calculated on the basis of the frequency of these error types (Waber & Holmes, 1986).

The Style score is intended to characterise drawing approach without reference to organisational structure, and was the only index examined in the current study as it represents the primary measure of planning in this scoring system. The Style rating is categorical with three categories specified: Part Oriented, Intermediate and Configurational. The Intermediate category is further broken down into Outer Configurational/Inner Part, and Outer Part/ Inner Configurational for Copy trials. The Style Score advances from Part-Orientated, through Intermediate, to Configurational over the course of the child's development. Children's drawings are therefore necessarily interpreted according to their developmental level. Each line segment is required to demonstrate good alignment and to be drawn continuously.

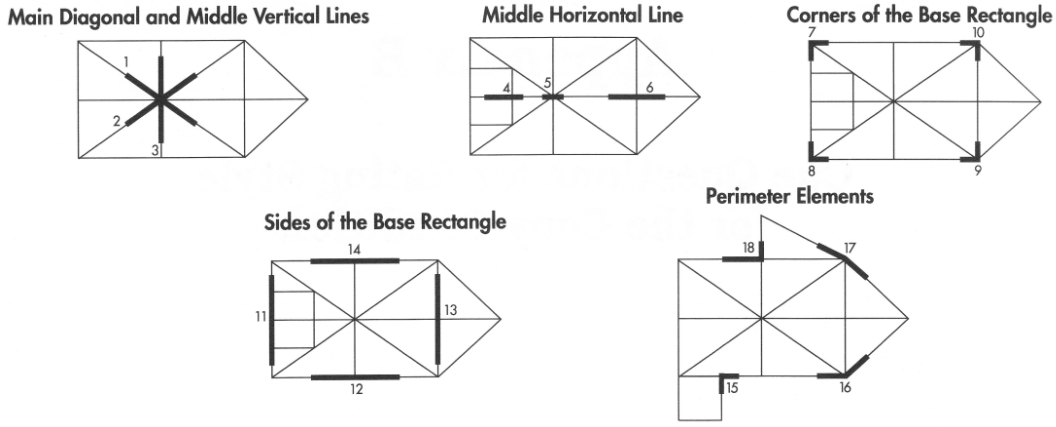

Figure 4. Scoring principle for continuity of drawing for line segments comprising Style Score [Source: Bernstein & Waber (1996) p. 78]

Identified features of the ROCF are weighted from 1-8 in the Style Score, with higher weights accorded to drawings considered to be more configural (Knight 2003b).

The authors reported that a random sample of 52 drawings was scored independently by two raters, revealing agreement in excess of 94% for accuracy, alignments and intersections, style junctures and errors. Inter-rater reliability was reported to be high for the organisation score across the Copy trial (0.94) and Recall trial (0.94), as was the Style score on the Copy trial (0.88) and Recall trial (0.85) (Waber & Holmes, 1986). Copy scores obtained using Waber and Holmes' scoring system were reported to be correlated with scores derived from the standard scoring system (r=0.60).

The DSS-ROCF has also been shown to discriminate normal children from clinical paediatric groups. The Organisation score for the Copy trial discriminated boys with Attention Deficit Hyperactivity Disoder ADHD from matched normal control subjects (Seidman et al, 1995), while a range of scores derived from this qualitative scoring system were found to discriminate normal controls from children with leukaemia (Waber et al, 1994) and children with traumatic brain injuries (Yeates et al, 2003).

The DSS-ROCF is one of a number of qualitative scoring systems developed for use in paediatric populations where, as with adults, there has been a strong interest in delineating performance parameters relating to poor planning/organisation versus poor recall (Akshoomoff & Stiles, 1995; Anderson et al, 2001).

Planning

It is difficult to understand process variables underlying performance on the ROCF without addressing the construct of planning. The ROCF remains widely used as a measure of planning, both in research settings (Poreh, 2006, Wilson & Batchelor, 2015) and clinically (Lezak et al, 2012; Weider et al, 2016). Other more recently developed measures of planning augment rather than replace the ROCF. In the neuropsychological literature, planning has been examined primarily in relation to problem solving tasks that require anticipation of events and consequences, additional to monitoring of goal attainment. The efficiency of individual search plans is examined in the Key Search Test, which comprises one of the subtests of an assessment battery, the Behavioural Assessment of the Dysexecutive Syndrome (BADS) (Wilson et al, 1996). Another subtest in this battery is the Zoo Map Test, which examines the sequence of steps taken to locate designated locations on a map, also assessing planning ability. As with a range of currently used measures of executive function, it is widely recognised that the demand characteristics of individual assessment measures can differ between individuals, both within and between clinical groups. There is increased awareness also of the many different ways performance can be compromised.

In the context of the ROCF, planning involves a number of stages. Initially it is important to perceive the figure's overall structure and component details, and subsequently identify how the figure elements relate to each other. Based on this analysis, formulation of the sequence order in which each element should be executed during construction of the drawing is undertaken. Compromised drawings

can then arise from poor implementation of a well planned approach or adequate implementation of a poorly planned approach. In the absence of a plan, it is more likely that copy productions of the ROCF will be executed in a fragmented manner (Knight, 2003b). Organisation is highly related to planning and these terms are often used interchangeably. It is reasonable to infer that a well organised drawing follows from a plan during the copy process. A poorly planned drawing is more likely to result in a disorganised and/or fragmented copy, especially if component details from the figure are drawn before structural features, increasing the likelihood of distortion or misplacement of these elements, and in turn, other aspects of the ROCF (Knight, 2003b). In the current study, the temporal aspects of planning were emphasised rather than the spatial aspects. Thus, drawing approach was characterised not in terms of the accuracy of the figure copy, but rather how individual elements of the ROCF were sequenced.

Tower tasks have been widely used to measure spatial planning and the capacity to follow rules during task execution. These rules typically discourage impulsivity and reinforce planning, thus contributing to enhanced task regulation. (Carey et al, 2008). The "Tower tasks" comprise a number of similar tests including the Tower of Hanoi (Anzai & Simon, 1979), The Tower of London (Shallice 1982) and the Delis-Kaplan Executive Function System (D-KEFS) Tower Test (Delis, Kaplan & Kramer, 2001). The Tower of London is the most frequently used test in this group and will be the sole focus of the present analysis.

Tower of London Test

The Tower of London Test (TOL) (Shallice 1982) was created as a measure of the Supervisory Attention System (SAS) proposed by Norman and Shallice (1986), the system responsible for the management of novel situations. This measure evaluates planning in the context of a means-ends analysis to solve a series of problems graded in task difficulty. It requires the movement of a starting configuration of three coloured beads on pegs, to match a target configuration, while making the minimum number of moves, and further anticipating and avoiding incorrect moves. Item difficulty is increased by increasing the minimum number of moves necessary to solve problems. Performance is measured by parameters including the number of problems solved, the total number of moves exceeding the minimum required for each goal state, response latencies and errors. Successful performance on this task requires analysis of the problem and planning the sequence of moves before the initiation of a response. While this measure is typically used to assess planning, other factors have also been identified to contribute to successful performance, including working memory (Carlin et al, 2000) and response inhibition (Phillips et al, 1999, Rainville et al, 2002).

The TOL has been widely used in clinical (Carey et al, 2008; Carlin et al, 2000; Rainville et al, 2002) and non-clinical (Kafer & Hunter, 1997; Morris et al, 1993; Schall et al, 2003) populations to assess the association of planning and frontal/executive function. In an early study, Shallice (1982) administered the Tower of London to 61 patients with unilateral lesions and 20 control subjects. The clinical group was subdivided into right/left posterior lesion and right/left frontal lesion groups. A significant interaction was observed between lesion location/laterality and

correct solutions achieved within one minute. Patients with left frontal lesions performed significantly more poorly on this measure than patients with left or right posterior lesions and patients with right frontal lesions. Owen et al (1990) used a slightly modified form of a computerised version of the TOL task developed by Morris et al (1988) in 26 patients who had undergone unilateral or bilateral frontal lobe surgery and 26 age-matched healthy controls. Clinical participants were found to demonstrate significantly longer time periods planning their responses, and used a significantly greater number of moves to solve individual problems. There were no significant differences reported between patients with left or right excisions, and there was no association identified between task performances and lesion size. Morris et al (1994) argue that the sample size used in this study may have been too small to demonstrate a laterality effect.

Mixed findings have been reported in clinical studies. Some studies have failed to find differences in performance on the TOL between clinical groups and controls. Cockburn (1995) compared performances on the TOL in 20 patients with severe, diffuse traumatic brain injury with that of 25 control subjects matched for age and education. Scores on the TOL did not discriminate between patient and control groups, with evidence of large differences in scores across subjects in both groups. Correlation analyses revealed the interaction of variables such as estimated premorbid intelligence and Post Traumatic Amnesia (PTA) duration on scores derived. Andres and Van Der Linden (2001) administered the TOL and another measure of executive function, the Hayling and Brixton Tests (Burgess & Shallice, 1997) to a mixed neurologic sample with identified frontal lesions, and a control group matched for age, gender and education. Using the number of moves taken to

solve problems as the performance index, the authors found that while patients with focal frontal lesions were slower on the TOL, their performance overall was comparable to control participants, as was their performance on the Hayling and Brixton Tests.

As with the ROCF, studies in this area highlight the difficulty of using overall test scores on the TOL, emphasising that these scores may not capture the underlying mechanisms of poor performance across patient groups (Carey et al, 2008). The Delis-Kaplan Executive Function System (D-KEFS) Tower Test (Delis et al, 2001) was used by Carey et al., (2008) to examine whether rule monitoring could discriminate the performance of 30 patients with probable Alzheimer's disease (AD) and 44 patients with Frontotemporal Dementia (FTD). These groups were compared with 27 healthy controls matched for age and education. Both patient groups exhibited difficulty with spatial planning on this task, demonstrating significantly poorer overall achievement scores relative to control participants. The mechanism underlying this difference was reported to be distinct for FTD patients who demonstrated more rule violation errors than both AD patients and control participants. The sensitivity of rule violation errors amongst FTD patients was, however, low, occurring in 50% of patients in this group. The authors identified that it was difficult to discern the mechanism underlying performance deficits for AD patients.

Comparisons have also been undertaken of TOL performances of patients with FTD and those with focal frontal lesions and normal controls (Carlin et al, 2000). Both clinical groups made more moves and demonstrated longer solution time latency relative to matched controls, but FTD patients made a higher frequency

of rule violations when compared with both control participants and patients with focal frontal lesions. Error patterns were noted to become more evident with increasing problem complexity. The authors concluded that several independent cognitive mechanisms contribute to the construct of planning as measured by the TOL and warn against defining a single causal factor (Carlin et al, 2000). Using a simplified version of the TOL in which simpler problems were embedded in more complex problems, Rainville et al., (2002) compared the performance of 17 patients with AD with 17 elderly control subjects. AD patients were found to perform more poorly relative to controls on problems demanding higher-level planning and, in contrast to findings reported by Carlin et al (2000) and Carey et al (2008), they made a greater percentage of rule breaking errors, which the authors inferred reflected an impaired capacity for self-monitoring (Rainville et al, 2002). These studies stress the importance of more comprehensively delineating the broad range of cognitive mechanisms which contribute to poor performances on the TOL

Psychometric investigations of the TOL have only revealed low internal consistency, with split-half reliability of 0.19 and Cronbach alpha of 0.25 (Humes et al, 1997). In their factor analysis, Kafer and Hunter (1997) concluded that the TOL has poor face validity as a measure of planning. There have also been wide variations across studies in terms of administration procedures, outcome measures, and item selection, thus making comparison across studies quite difficult (Kaller et al, 2004; Krikorian, Bartok & Gay, 1994).

Given the inherent limitations of the Tower Test, it was decided that the ROCF would potentially represent a better measure of planning, with sequence order reflecting the integrity of plan execution.

Organisational Strategies and Recall of the Rey Osterrieth Complex Figure

The ROCF is ideally suited to characterise patterns of visual memory impairment given the opportunity to observe the organisational strategies adopted during copy administration of the figure. Assuming that recall of the ROCF is influenced by planning and organisation strategies, poor recall performance can then reflect poor initial organisation of the figure, or loss of information organised adequately. It is well documented that patients might produce reasonably accurate copies, but adopt a disorganised approach. Generally, and as detailed below, more efficient strategies are associated with better recall relative to fragmented or disorganised strategies. Recognition of process variables therefore represents an integral part of evaluating recall performance on the ROCF. The development of qualitative scoring systems by which these process variables can be more readily characterised represents an important advance in this area.

Bennett-Levy (1984) was one of the first researchers to examine process variables and their association with recall on the ROCF. In a sample of normal adults aged 17-49, copy strategy measures denoted by Symmetry, Good Continuation and Strategy Total scores were examined alongside demographic variables including age and estimated IQ in relation to their ability to predict copy accuracy and recall performance on the ROCF. Analyses revealed that strategy

measures, copy score and age were the best predictors of recall performance. Given the degree of variance accounted for by copying strategy and age, a regression equation was derived to predict delayed recall on the basis of these variables (Bennett-Levy, 1984). These findings are yet to be replicated in a clinical sample.

Shorr, Delis & Massman (1992) proposed that recall of visuospatial material should be more strongly mediated by the process by which the material is encoded during learning, rather than copy accuracy. They contest that recall of complex visuospatial material is ideally enhanced if organised into meaningful perceptual units during encoding. Scores derived from the qualitative scoring system developed by these authors were examined in this context using archival neuropsychological data for a mixed neurological sample aged 23-83 years. A history of alcohol abuse was reported in 22% of this clinical sample. Performance measures used in the analysis involved Copy Accuracy (derived from standard scoring system for the ROCF), Perceptual Cluster Ratio, Immediate Recall Accuracy, Encoding, Delayed Recall Accuracy and Savings score (delayed recall accuracy relative to immediate recall accuracy). A significant correlation was reported between the Perceptual Cluster Ratio and the Encoding score, which was stronger than the correlation between the Copy Accuracy score and the Encoding score. It was further reported that the Perceptual Cluster Ratio made a significant, unique contribution to predicting the Encoding score, in contrast to Copy Accuracy which was not predictive of performance on the Encoding score. Both the Perceptual Cluster Ratio and Copy Accuracy were not significantly correlated with the Savings score, which the authors inferred reflected the different cognitive mechanisms underlying these separate indices (Shorr et al, 1992).

Grossman et al (1993) adopted a modified version of the standard scoring system for the ROCF to examine performances on copy, 1-minute and 5-minute recall trials in patients with Parkinson's disease (without dementia) compared with normal controls. Structural elements and detail elements of the ROCF were separately characterised by the authors, drawing on Binder's (1982) description of configural elements of the ROCF. Parkinson's patients performed more poorly across both copy and recall trials. They were observed not to construct the main structural elements in an organised manner at the beginning of their drawings like control subjects, but rather demonstrated a more fragmented copy of structural elements. Parkinson's patients also failed to recall a greater number of structural elements relative to control subjects. Impaired organisational strategies were found to account for most of the variance in both copy and recall trials in this clinical group. This is consistent with findings reported in a study of impaired memory performance in patients with ruptured and repaired anterior communicating artery aneurysm, where provision of an organisation strategy for encoding details of the ROCF enhanced immediate recall and consolidation of this material over a 30-minute delay (Diamond et al, 1997). The authors acknowledged that a potential weakness of their study was the absence of a control group receiving repeated exposure of the ROCF without provision of an organisation strategy, thereby limiting inferences regarding the primary mechanism for improved recall.

The effect of copy approach on recall of the ROCF has been examined in a non-clinical, well educated, elderly sample aged 70 to 93 years (Chiulli et al,1995). The use of a "configural" approach (large rectangle drawn first on copy) was

reported to be associated with stronger performances across Copy, Immediate, and 30-minute Delayed recall for all age groups.

The qualitative aspects of recall performance on the ROCF have been widely examined in individuals with Obsessive Compulsive Disorder (OCD) (Savage et al, 1999; 2000; Deckersbach et al, 2000). Organisational strategy was based on the configural elements characterised by Binder (1982), with the addition of the central rectangle, which was accorded extra weight in order to reflect its importance in the organisation of the figure. When compared to a matched group of healthy control subjects, group differences were evident on Immediate recall, attributed to the impaired ability of OCD patients to organise nonverbal information in the context of meaningful organisational units during encoding. Once learned, however, this information was well retained, leading the authors to conclude that observed memory problems were secondary to impaired executive strategies during learning (Savage et al 1999). These findings were replicated in subsequent research where "impaired strategic processing" (p. 147) was again demonstrated in the performance of OCD patients on the ROCF relative to healthy controls (Savage et al 2000). Parallel findings were also observed on a verbal learning measure (CAVLT) in the clinical group in this same study. Significantly poorer performances on immediate recall of the ROCF were also reported in a sample of 35 OCD patients relative to 33 healthy controls, which were not, however, matched for age (Penades et al, 2005). Adopting Savage et al's (1999) qualitative scoring system, OCD patients were found to demonstrate poor organisational strategies during copy administration of the ROCF relative to controls, and performed significantly worse on Immediate recall and across a range of executive measures. Regression modeling identified copy organisation as a

strong predictor of recall performance, which the authors argued lent support to the assertion that memory impairment in OCD is likely to be secondary to executive dysfunction (Penades et al, 2005).

The enhanced effect on recall of a strategy for organising the ROCF has also been widely demonstrated in other clinical groups. Organisational strategies and the influence on accuracy of figure construction has been examined in groups of patients with a history of alcohol dependence (Sullivan et al,1992). Recall accuracy was reported to be contributed by the use of organised copying strategies, highlighting the significance of organisational approach during encoding. Dawson and Grant (2000) also found that the organisational approach to the ROCF accounted for a large proportion of the variance in immediate and delayed recall trials in a group of recently detoxified alcoholics. These authors used the Boston Qualitative Scoring System (BQSS) to capture copying strategy, finding that compared with long term abstinent patients and normal controls, recently detoxified alcoholic patients demonstrated more impaired organisation, perceptual clustering and constructional accuracy, which negatively impacted subsequent recall. The authors concluded that their study findings provided supportive evidence for the notion that better organisation of spatial information during encoding enhances accuracy of recall.

Kixmiller et al (2000) developed an organisational scoring system for the ROCF in their broader assessment of visual-perceptual accuracy, organisation and memory performances across three clinical conditions: (1) Alcoholic Korsakoff's syndrome, (2) Medial temporal damage; and (3) Anterior communicating artery (ACoA) aneurysm rupture and/or repair. Korsakoff syndrome patients demonstrated

significantly poorer copy accuracy than the other two clinical groups, and a control group. Significant differences were also evident between ACoA patients and the control group. Both of these clinical groups demonstrated a more disorganised approach than patients in the medial temporal group and controls. Korsakoff patients showed a greater tendency to omit figural details, with further distortion of the spatial relationship of figural items. ACoA patients were more likely to approach their copy of the ROCF in an impulsive manner, manifest by poor planning and poor attention to detail. The primary area of weakness for the medial temporal patients involved errors of omission. The control group outperformed all three clinical groups on immediate and delayed recall trials. Korsakoff patients performed most poorly of the three groups at immediate and delayed recall.

Lange et al (2000) investigated the influence of organisational strategy during encoding of the ROCF on recall in a clinical sample of 37 stroke patients. Patients were initially divided into two groups, Right CVA or Left CVA, and were later re-classified as either Cortical CVA or Subcortical CVA. Copy, Immediate and Delayed recall trials of the ROCF were evaluated. Copy accuracy and recall scores were derived using the standard scoring system. A modification of Binder's (1982) Configural Score was used to define organisational strategy such that the number lines taken to draw the "three key elements" of the ROCF - (i) central rectangle, (ii) vertical and horizontal lines, and (iii) diagonals - was noted. These elements could be drawn contiguously and in any order. Findings were reported to show that the Right CVA group were more likely to approach their copy of the ROCF in a fragmented, disorganised manner and demonstrated poorer constructional accuracy relative to the left CVA group. Organisational strategy scores were significantly related to copy

accuracy and Immediate and Delayed recall across both clinical groups. Recall scores were not significantly different between groups. When the group was then subdivided into Cortical CVA and Subcortical CVA, patients in the Cortical CVA group demonstrated significant impairments in organisational strategy and copy accuracy relative to the Subcortical group. Again, there were no group differences in Immediate or Delayed recall of the ROCF. The authors concluded that organisational strategy during encoding of the ROCF was associated with improved copy accuracy and enhanced recall (Lange et al, 2000).

Westervelt, Somerville, Tremont and Stern (2000) investigated the influence of organisation strategies during copy of the ROCF on both immediate and delayed recall of the figure in a mixed neurological sample. The sample was sub-divided into High- and Low- Organisation groups on the basis of Organisation Scores derived from the BQSS (Stern et al, 1999). Significant group differences were observed across both Immediate and Delayed recall of the ROCF. Parallel findings were reported by Temple and colleagues (2006) in a mixed sample of 193 neurological outpatients where organisation strategies during copy of the ROCF were observed to be related to subsequent recall. The Organisation Score from the BQQS was predictive of recall for both the ROCF and the Visual Reproduction subtest of the Wechsler Memory Scale-Third Edition (WMS-III), beyond the variance accounted for by other demographic variables, MMSE score, and a range of measures of executive function. Consideration was given to the possibility that the observed relationship between organisation measures and recall of the ROCF could reflect method variance (comparing two measures of the same task), but this was deemed unlikely due to the significant relationship between the organisation indices of the BQSS and Visual

Reproduction recall, above and beyond the executive measures. Executive measures were predictive of Visual Reproduction performance variables after demographic variables and the MMSE scores were taken into account.

These studies collectively suggest that patients with executive dysfunction are more likely to demonstrate poor organisation of the ROCF. This has prompted further research addressing whether individuals with adaptive/executive deficits benefit from assistance with strategies targeting planning and organisation. Buhlmann and colleagues (2006) investigated the efficacy of a brief cognitive retraining program for organisational deficits in a patient group with OCD. Both normal controls and OCD patients improved their organisational skills with enhanced recall of the ROCF, however, OCD patients undergoing training did not improve more than OCD patients not undergoing training. There was no consideration given to the possibility that repeated administration of the ROCF in a short time interval may have contributed to practice effects rather than genuine improvement. The authors concluded that OCD patients demonstrate difficulty employing effective organisation strategies when initially copying the ROCF, but improve their strategy when provided another opportunity to copy the figure, without specific cognitive retraining. It was postulated on that basis that OCD patients experience difficulty relating to the spontaneous initiation of organisational strategies during encoding, while their ability to implement these strategies remains preserved. These findings have been challenged by other researchers who have reported only a weak relationship between visual memory impairment and organisational strategies used during encoding (Shin et al, 2004). Shin et al (2010) examined the influence of organisational skills on memory deficits in drug-naïve patients with OCD. They reported a minimal effect of

organisational strategies derived from the Boston Qualitative Scoring System (BQSS) (Stern et al, 1999) on Immediate and Delayed recall of the ROCF in this sample, arguing that the mechanism underlying memory dysfunction in OCD remains unclear. Clearly, further research is required in this area to more definitively establish the mechanisms underlying poor recall of the ROCF in clinical groups demonstrating executive deficits.

Research on the association between copy approach and recall of the ROCF highlights the importance of ongoing investigation of process variables in promoting a more refined and accurate analysis of compromised performance on the ROCF.

Process Analysis

Edith Kaplan (1988) identified that: The complexity of the Rey Complex Figure: "lends itself to a process analysis" (p.156). Kaplan (1983, 1988**)** hypothesised that the effectiveness of a learning strategy is increased if use is made of the organisational features inherent in the stimuli. Attempts have been made by clinicians to quantify the process by which individuals perform neuropsychological assessment measures in an attempt to better understand underlying cognitive processes. The Quantified Process Approach (Poreh, 2000; Poreh 2006) has evolved from the "Boston Process Approach" in which qualitative aspects of the patient's performance are quantified and subjected to statistical analysis, and where testing of clinical limits is operationally defined. Standardised administration and scoring of commonly used neuropsychological tests is generally not adhered to. The Boston Process Approach has been criticised for its total reliance on case studies and it has

been argued that clinical assertions thus obtained cannot be subjected to empirical enquiry (Erickson, 1995; Milberg & Hebben, 2006). The Quantified Process Approach differs from the Boston Process Approach in that standardised administration and scoring are maintained, thus allowing for replicability of findings. This approach further advocates nomothetic analysis of data additional to implementation of robust criterion measures.

Quantified Process Approach

In his preface to Amir Poreh's book: "The Quantified Process Approach to Neuropsychological Assessment", David Loring (2006) writes: "Being able to measure constructs and derive new process measures by combining or developing information from existing procedures … is the necessary first step to evaluate their clinical and scientific merit" (xxiii).

The Quantified Process Approach uses three major methodologies: (1) The Satellite Testing Paradigm, where new tasks are developed to complement existing tests in order to better characterise test performance (e.g., WAIS-R NI; Kaplan, Fein, Morris & Delis, 1991); (2) The Composition Paradigm, where data collected for a given test, but not previously examined, is compiled and analysed, resulting in the generation of new empirical indices assumed to characterise the process and strategies underlying individual test performances. The aforementioned qualitative scoring systems for the ROCF represent new indices using the composition approach; and (3) the Decomposition Paradigm, where the relationship between test items of a given measures are investigated based on underlying facets, leading to the development of new subscores. This is exemplified by Poreh and Shye's (1998) use

of theoretically based decomposition models of the ROCF and their subsequent validation of derived indices using statistical measures examining the similarity between variables.

The Satellite Testing Paradigm essentially involves the creation of new tasks to supplement existing tests. One difficulty with many of the satellite tests that have been developed is the lack of clear guidelines relating to how these tests should be used and the absence of research supporting their added value relative to information provided by traditional testing. Another criticism relates to concern over the effect of administering satellite tests on the psychometric properties of the tests they supplement. Satellite paradigms have also come under scrutiny since they lengthen the assessment process. This is potentially problematic with the push towards cost-effective neuropsychological evaluations.

The Composition and Decomposition Paradigms avoid the aforementioned problems, but have developed in a non-systematic fashion. This is exemplified by the numerous qualitative scoring systems for the ROCF where there is inconsistent reporting of reliability and validity of measures thus derived. Of particular importance is demonstrating incremental validity, that is, demonstrating that quantitative measures obtained from these paradigms assist is prediction beyond that obtained in existing clinical practice.

The qualitative scoring systems examined thus far have been developed to quantify approaches to reproducing the ROCF. Osterrieth's (1944) system adopted categorical scales. The remaining qualitative scoring systems

reviewed have adopted numerical scales (Bennet-Levy, 1984; Binder, 1982; Savage et al 1999, Shorr et al, 1992; Waber & Holmes, 1986). Troyer and Wishart's (1997) review of these systems revealed that only four systems demonstrated approximately normal distributions with moderate variability, which Troyer and Wishart argue are requisites for enabling group comparisons and identifying changes over time.

Poreh (2006b) highlights that while process measures for the ROCF and other complex figures should ideally be theory driven, normal individuals don't always use strategies deemed to be optimal on theoretical grounds. A shift in thinking is suggested from theoretical assumptions about optimal performance on the ROCF to formulations on how normal individuals approach this task, thus enabling empirically derived indices based on the performance of a normal reference group. This empirical approach to providing quantified process measures for the ROCF is yet to be established.

There is a dearth of research examining qualitative performance on the ROCF in healthy individuals where it has been largely assumed that the central rectangle represents the primary organisational feature. Normal populations have been used as control groups in qualitative analyses of the ROCF (Osterrieth 1944, Binder 1982; Casey et al 1991; Ska & Nespoulous 1988, Visser 1973; Waber & Holmes 1985; Waber & Holmes 1986). Individual differences exist, however, in how normal individuals approach their copy of the ROCF.

Within the framework of Osterrieth's seven reproduction types, reproduction types I and II were identified as the superior approach adopted by most

normal individuals (Osterrieth 1944, cited in Corwin & Bylsma 1993). Eighty-three percent of the control group were reported to adopt these reproduction types which were referenced to the central rectangle. Type IV, indicating a piecemeal approach, was noted to be most indicative of cognitive impairment and was evident in 15% of the normal sample (Lezak et al, 2012).

Recent studies have identified that Reproduction Types III and IV are more prevalent in normal individuals than originally assumed. Maillet (1992) compared Osterrieth's Reproduction Types and one of the measures derived from Bennett-Levy's scoring system, Copying Strategy, in a normal sample aged 16-80. Only 59% of adults were found to adopt Reproduction Types I and II, while 26.7% of normal subjects adopted Reproduction Type IV. Ska and Nespoulous (1988) examined Osterrieth's Reproduction Types in relation to recall of the ROCF in a normal sample of 150 subjects divided into 5 age bands. Across the first two age bands spanning 20-50 years, there was almost equal representation of individuals adopting Reproduction Types I and II, as there were Type IV. Similarly, in the age band 55-64, only 53% of individuals adopted Reproduction Types I and II, and 44% adopted Reproduction Type IV, paralleling findings for age band 65-74, with 54% demonstrating Reproduction Types I and II and 41% demonstrating Reproduction Type IV. In the oldest age band, 75-84 years, only 30% of individuals adopted Reproduction Types I and II, while 60% adopted Reproduction Type IV (Ska & Nespoulous, 1988).

In a more recent study of undergraduate students, the Q-Score from Bylsma's (2008) qualitative scoring system was used to evaluate reproductions of the

ROCF (Wilson & Batchelor 2015). Substantial variations in performances were evident in this normal, healthy sample, with evidence of fragmented, piecemeal encoding of the stimulus on copy. The authors report that only 53% of the subject sample commenced their copy of the ROCF with the central rectangle drawn continuously. These findings underscore the considerable heterogeneity in normal adults when constructing their copy the ROCF, further highlighting difficulties inherent in theoretically driven measures of suboptimal performance on the ROCF and the potential merits of empirically derived measures.

The diverse nature of featured qualitative scoring methods for the ROCF reflects the longstanding divide in opinion regarding the target of analysis when evaluating suboptimal performance. Poreh's quantified process approach (Poreh, 2006) provides a framework within which empirically derived process measures based on the performance of a normal reference group can be derived. Further investigation of the utility of empirical approaches over theoretically based scoring systems is desirable.

Aims

The aim of the current study was to develop quantified process measures for the ROCF relating to planning, which are not based on theoretical assumptions regarding optimal performance but are instead empirically derived, determined on the basis of the performance of a normal reference group. In order to establish the clinical utility of variables thus derived, they were compared to a subset of the qualitative scoring systems examined by Troyer and Wishart (1997). The Organisational Quality Score developed by Hamby et al (1993) and Stern et al's

(1994) Fragmentation and Planning Scores were excluded as these systems required reference to accuracy of ROCF reproductions which was not considered in the planning process measures under study. In line with Troyer and Wishart's analysis, consideration was given to psychometric parameters such as distribution of scores, specifically whether the process variables produced scores which were normally distributed and demonstrated variability between subjects. Consideration was also given to whether measures thus obtained could discriminate between normal and clinical subjects. Quantified process measures were also evaluated in relation to their ability to predict recall.

Hypotheses

It was predicted that:

(1) Quantified process measures would produce scores that were normally distributed and which varied between subjects. It was predicted that visual inspection of score distributions would appear normal and be validated by skewness and kurtosis statistics.

(2) Novel quantified process measures would demonstrate at least moderate test-retest reliability

(3) Novel quantified process measures would discriminate between normal controls and a heterogeneous clinical sample. It was predicted that logistic regression with group as the dependent variable would reveal significant effects for novel quantified process measures.

(4) Novel quantified processing measures would be related to recall of the ROCF, with at least moderate correlations between quantified process measures and recall of the ROCF.

Chapter 2

Method

An existing clinical data set was used in the current study, for which ethics approval had been previously granted from Macquarie University Ethics Review Committee (Human Research) (Ethics Ref: 5201001467).

Participants

Normal Control Sample

The data used for analysis in the current study had been collected by a Co-Investigator, Dr Jamie Berry, as part of a larger study. Normal control participants were recruited through personal contacts, acquaintances and colleagues. No incentives were provided for participation in the study. Participants were advised that individual assessment sessions would be of approximately 30 minutes duration. During each assessment session, the Co-Investigator administered the Rey-Osterrieth Complex Figure Test, additional to the Bells Test, Clock Drawing Test, and the Five-Point Test. The control sample comprised 49 participants, of which 16 were male and 33 female. The average age was 46.9 years (SD: 15.3 years, Range: 18-82 years), and the average level of education was 13.0 years (SD: 2.76 years, Range: 9-17 years).

Exclusion criteria were adopted as follows:

(1) Poor proficiency with the English language such that comprehension of task instructions might be adversely influenced;
(2) Current involvement in a treatment program for drug and/or alcohol dependence;

(3) Current consumption of more than three standard alcoholic beverages more than two nights weekly;

(4) Currently seeing a medical practitioner or other health professional for problems with memory or thinking;

(5) Difficulties with upper limb function limiting the ability to write or draw;

(6) Loss of consciousness (>1 hour) following a traumatic brain injury; and

(7) Past history of a neurological or psychiatric disorder.

The normative sample was divided into two groups for the purposes of development and validation of the quantified process measures. The quantified process measures were created using Split Group 1. Split Group 2 was used as the validation group for the quantified process measures, and became the reference group for statistical analyses of these indices. This group was subsequently referred to as the Split Normal group. The demographic characteristics of the entire normative sample and the split groups are detailed in Table 1.

Table 1. *Demographic Data for the Full Normal Group and Split Normal Groups*

Variable	Full Normal	Split Group 1	Split Group 2
N	49	24	25
Mean Age (SD)	46.9 (15.3)	46.9 (13.9)	46.9 (16.8)
Mean Years of Education (SD)	13.0 (2.76)	13.0 (2.92)	13.1 (2.66)
Proportion Males: Females	16:33	5:19	11:14

The average age of Split Group 1 did not differ significantly from the average age of Split Group 2, $t(47) = -0.09$, $p = 0.371$. The average education level of Split Group 1 did not differ significantly from the average education level of Split Group 2, $t(47) =$

0.10, p = 0.639. The proportion of males to females did not differ significantly between the split groups, X (1, N=49) = 2.988, p= 0.084.

A second Normal group was used to quantitatively delineate measures of variability for qualitative scoring systems under review for a younger control sample. This sample was taken from a study of university students undertaken by Wilson and Batchelor (2015), where the ROCF data were collected by Nikki Wilson, the chief investigator of that study. A subset of the de-identified ROCF data was accessed and used for analysis in the current study, with approval having been previously granted by the Macquarie University Ethics Review Committee (Human Research) (Ethics Ref: 5201400311). Six subjects were excluded because demographic details were not clearly stated on source documents, resulting in a final number of 51 participants. This sample was subsequently referred to as the Wilson Normal group. The exclusion criteria for that study were as follows:

(1) Poor proficiency with English;
(2) Past history of cardiac or blood pressure problems; and
(3) Past history of clinical depression, anxiety disorders, Obsessive Compulsive Disorder (OCD), Body Dysmorphic Disorder (BDD), or Attention Deficit Hyperactivity Disorder (ADHD).

The Wilson Normal group used in the current study comprised 13 males and 38 females. The proportion of males to females did not differ significantly between the Full Normal group (N=49) and the Wilson Normal group, X (1, N= 100) = 0.623, p= 0.43. The average age of the Wilson Normal group was 19.96 years (SD: 3.46 years,

Range: 18-34 years), which differed significantly from the Full Normal group, $t(98) = 12.05$, $p < 0.001$. The average level of education for the Wilson Normal Group was 12.2 years (SD:0.81), which differed significantly from that of the Full Normal group, $t(98) = 2.011$, $p<0.01$.

Clinical Sample

Clinical participants comprised consecutively referred patients to St Joseph's Hospital Medical Rehabilitation Neuropsychology Service between September 2006 and April 2009. A subset of the original sample of 101 patients who had also been administered the ROCF as part of a broader neuropsychological assessment battery were selected. Administration was undertaken either by a Co-Investigator (Dr Jamie Berry) or intern neuropsychologists on placement during that period**.** The average age of the Clinical group was 52.1 years (SD=13.1, Range: 17-78 years), which did not differ significantly from the Full Normal group, $t(99) = -1.819$, $p = 0.132$, or the Split Normal group, $t(75) = 1.475$, $p = 0.144$. The average education level of the Clinical group was 11.4 years (SD=3.33), which differed significantly from the Full Normal group as revealed in an ANOVA comparing education for the normal groups and the clinical group, $F(2, 149) = 5.11$, $p=0.007$, and from the Split Normal group, $t(75) = -2.175$, $p = 0.033$. The Clinical group comprised 30 males and 22 females. The proportion of males to females differed significantly between the Full Normal and the Clinical group, $X (1, N=101) = 6.37$, $p=0.012$, but not between the Clinical group and the Split Normal group, $X(1, N=77) = 1.271$, $p = 0.259$. Participants were excluded if the ROCF copy drawings were poorly registered on the digital notepad (Digimemo) used to record ROCF reproductions. Among the Clinical group, diagnoses included right CVA (n=10), left CVA (n=9), bilateral CVA (n=10),

subarachnoid haemorrhage (SAH) (n=6), neurodegenerative condition (n=5), neoplasm (n=6), Multiple Sclerosis (n=1), alcohol related acquired brain impairment (n=2), hypoxic brain injury (n=2), and cerebral infection (n=1).

Table 2. *Demographic Data for Normal Groups and Clinical Group*

Variable	Split Normal	Wilson Group	Clinical Group
N	25	51	52
Mean Age (SD)	46.9 (16.8)	19.96 (3.46)	52.1 (13.05)
Mean Years of Education (SD)	13.1 (2.66)	12.2 (0.81)	11.4 (3.33)
Proportion Males: Females	11:14	13:38	30:22

Re-test Sample

Sixteen percent (N=8) of the Full Normal group were tested on a second occasion. The sample size for this group was small because a number of cases who had undergone re-testing for research purposes were excluded due to poor capture of their ROCF copy drawings on the digital notepad. The mean re-test interval was 155 days, (Range: 5-464 days, median: 70 days; Interquartile range: 65-217 days). The average age of the Re-Test group was 37.5 years (SD: 13.21, Range: 26-54). The difference in age between the Full Normal group and the Re-test Group only just failed to reach significance, $t(47 = 1.963, p = 0.056)$. The average level of education for the Re-test group was 13.88 years (SD: 2.295, Range: 9-16), which did not differ significantly from the Full Normal group, $t(47 = -0.933, p = 0.356)$. The proportion of males to females did not differ significantly between the Full Normal group and the Re-test group, $X (1, N=57) = 0.102, p= 0.749$.

Table 3. *Demographic Data for Re-Test Sample*

Variable	Re-Test Group
N	8
Mean Age (SD)	37.50 (13.21)
Mean Years of Education (SD)	13.88 (2.30)
Proportion Males: Females	3:5

Measures

The primary measure of interest was the Rey-Osterrieth Complex Figure Test (ROCF). Copy scores based on the standard scoring system were derived in the current study, additional to four quantified process measures. Derivation of these process measures is detailed below, based on the 18 elements detailed in Figure 5. A number of elements had sub-components (e.g., horizontal midline comprised 4a, 4b, 4c) thus enabling determinations around elements which were drawn continuously versus elements which were drawn in a fragmented manner.

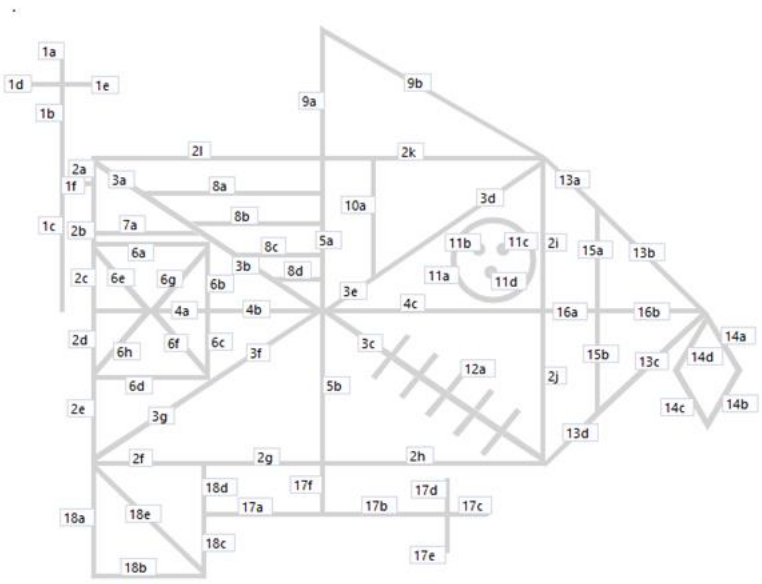

Figure 5. Rey-Osterrieth Complex Figure Scoring Template

Berry Approach Completed (BAC) – The BAC Index was derived by totaling the proportions corresponding to completion order for each of the 18 identified elements from the normal sample relative to the subject's completion order. These values are provided in Table 10, Appendix 2. The elements of the ROCF were ranked according to the order in which they were completed by the subject. Element order was absolute across all of the four novel process indices. The first completed element was ranked 1, the second completed element was ranked 2, and so on. For each element ranking, the proportion of the normal sample completing this element in this order was identified. The sum of these ranking proportions was totaled to derive the index for individual subjects.

Berry Approach Completed Weighted (BAC-W). The BAC-W Index was developed to weigh more heavily earlier item choices completed. The first result for items comprising the BAC Index was multiplied by 18, the second by 17, the third by 16, and so on. In the case of 18 elements having been completed, the last result was multiplied by 1. These values are provided in Table 11, Appendix 2.

Berry Approach Attempted (BAA) – The BAA Index was derived by summing the proportions of the order in which elements were attempted from the normal sample corresponding to the subject's element attempted order (Table 12, Appendix 2). Again, the elements of the ROCF were ranked according to the order in which they were attempted. For individual element rankings, the proportion of the normal sample attempting this element in the same order was identified, and the ranking proportions totaled to derive the Index.

It was deemed important to assess both completed and attempted elements of the ROCF in two separate indices given that cognitive deficits such as constructional impairments and neglect can underlie performance deficits on the ROCF. This was especially important in the current study where the clinical group largely comprised patients who had suffered a stroke.

Berry Approach Attempted Weighted (BAA-W). The BAA-W Index was developed to place emphasis on earlier item choices attempted. The first result for items comprising the BAA Index was multiplied by 18, the second by 17, the third by 16, and so on (Table 13, Appendix 2). In cases where all 18 elements were attempted, the last result was multiplied by 1.

The process of weighting attempted items was used to address the difficulty of incomplete figures, specifically to reduce the effect on the derived index of incomplete figures. This way, the index ideally reflected the approach taken, not the outcome.

High scores were proposed to denote the proximity of individual copy approaches to the strategy generally adopted by the normal reference group for the current study.

Each of the eight qualitative scoring systems examined thus far was programmed for comparison against the derived quantified processing measures of the ROCF as follows:

Osterrieth (1944) – Individual drawings were categorised according to identified Reproduction Types, though categories IV-VII were grouped together given that figure accuracy was not examined in the current study.

Visser (1973) – A Total Score comprised of each of three category scores for Interruptions, Omissions and Sequence Items was derived. There was no theoretical maximum score.

Binder (1982) – A Configural Score ranging from 0-5 was calculated, with credit given when each of the five elementary units identified from the ROCF was drawn as continuous line segments.

Bennett-Levy (1992) – A Strategy Total Score was calculated by combining the Good Continuation Score (Range: 0-18) and the Symmetry Score (Range: 0-18).

Shorr et al (1995) – A Perceptual Cluster Ratio was calculated by dividing the Perceptual Cluster Score (Range: 0-20) by the total number of junctures present (Range: 0-20), and multiplying this figure by 100.

Bylsma (2008) – The Q-Score was calculated on the basis of 13 defined unit points and 9 order points (Range: 0-24).

Savage et al (1999) – A Total Score was calculated based on the presence of the configural elements defined by Binder (1982) (Range: 0-4), with the addition of the

central rectangle which was weighted more heavily (2-points), providing a Total Score of 0-6.

Waber & Holmes (1985) – The Style Score was used from the DSS-ROCF, where credit was given for identified lines having been drawn continuously (Range: 0-18).

Visser's (1973) qualitative system was the only index where high scores reflected poorer performances.

For ease of reference the name of the first author was used to label existing scoring systems across all analyses.

Procedure

Ethical approval for the study in which data collection was completed was granted by the Macquarie University Ethics Review Committee (Human Research) (Ethics Ref: 5201600701). The study was undertaken in accord with the Australian Code for the Responsible Conduct of Research (NHMRC, ARC & UA, 2007).

All study participants undertook the ROCF on a digital notepad (Asus RIF, Intel Core Duo Processor T7200@2.0 GHz 1 Gig RAM, Embedded Intel graphic, 12" screen size, 1280x800 landscape mode screen resolution). The electronic administration was constructed to precisely replicate the paper and pencil version of the ROCF and standard instructions were administered. Subjects used a pen to execute their drawings on a sheet of paper placed on the tablet, for both copy and 30-minute delayed recall administration.

Data Scoring

The order of copy execution was recorded on flowcharts for each subject. This information was used to document how individual elements of the ROCF were drawn, both in terms of line direction and sequence of defined elements. Flowchart information was manually entered into an online SQL database from which quantified process measures were calculated. Individual elements of the ROCF (n=18) were broken down as detailed in Figure 5. This breakdown was decided upon so that the scoring criteria for each of the eight qualitative scoring systems under examination could be met.

The study investigator and a co-investigator undertook flowchart analyses on a random sample of 50% of copy drawings from each of the Full Normal, Clinical and Re-test groups. Both investigators were highly experienced in the administration and interpretation of the ROCF. Inter-rater reliability between scorers was 99.9% for the Normal group, 98% for the Clinical group, and 98.5% for the Re-test group. A random sample of protocols from each of the study groups (10%) was also scored manually by the author to validate the scoring algorithms generated for each of the existing qualitative scoring systems.

Data Analysis

SPSS was used to conduct all statistical analyses.

Chapter 3

Results

Distribution of Scores and Variability of Scores Between Subjects

It was predicted that quantified process measures would produce scores that are normally distributed, evident both by visual inspection of score distributions and through skewness and kurtosis statistics. The Split Normal group used for descriptive analyses of the quantified process measures (N=25) was much smaller than that used for descriptive analyses of existing qualitative scoring systems where the Full Normal sample of 49 subjects was used. However, when score distributions for both quantified processing measures and existing qualitative scoring systems using this smaller sample were compared to the larger sample, a similar pattern of findings was evident as is demonstrated in Table 4. The histograms for these analyses are presented below.

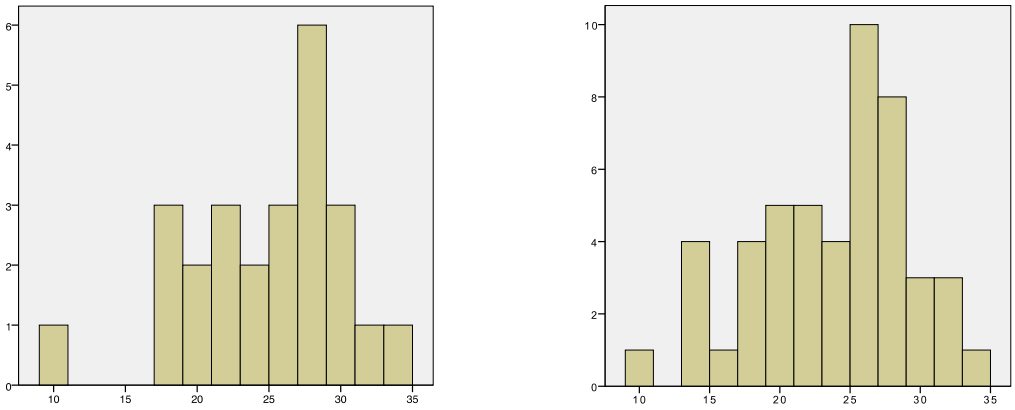

Figure 6. Score distribution for Bennett-Levy for the Split Normal group (N=25) (left) and Full Normal group (N=49) (right)

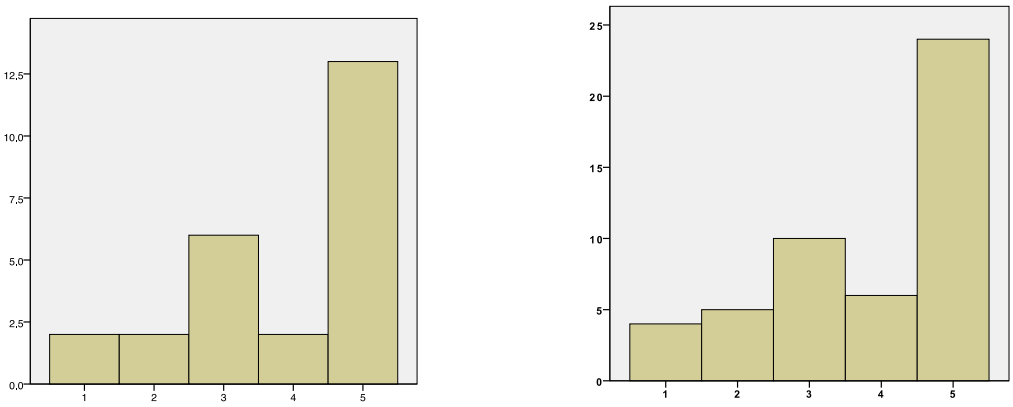

Figure 7. Score distribution for Binder for the Split Normal group (N=25) (left) and Full Normal group (N=49) (right)

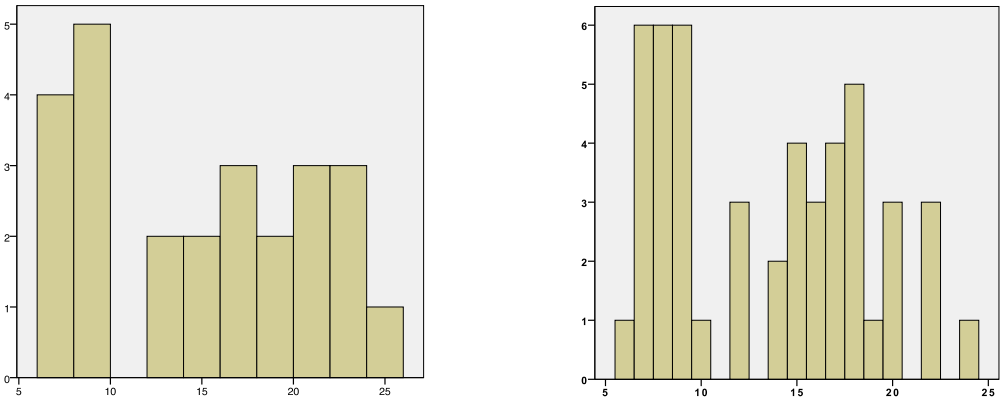

Figure 8. Score distribution for Bylsma for the Split Normal group (N=25) (left) and Full Normal group (N=49) (right)

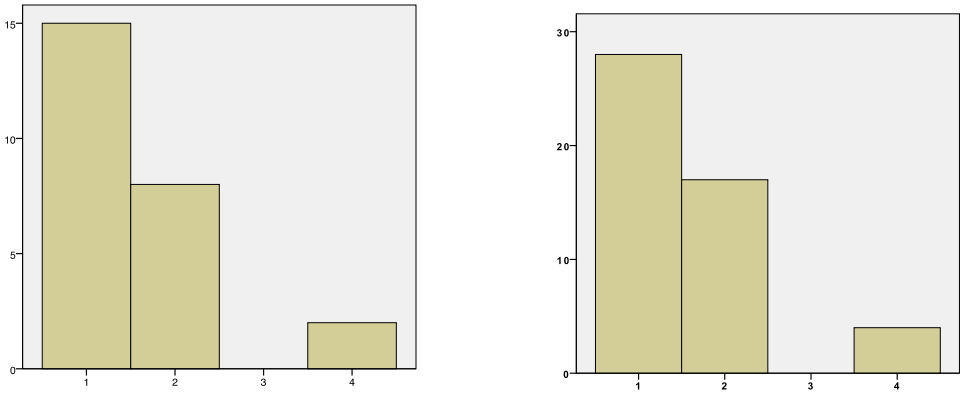

Figure 9. Score distribution for Osterrrieth for the Split Normal group (N=25) (left) and Full Normal group (N=49) (right)

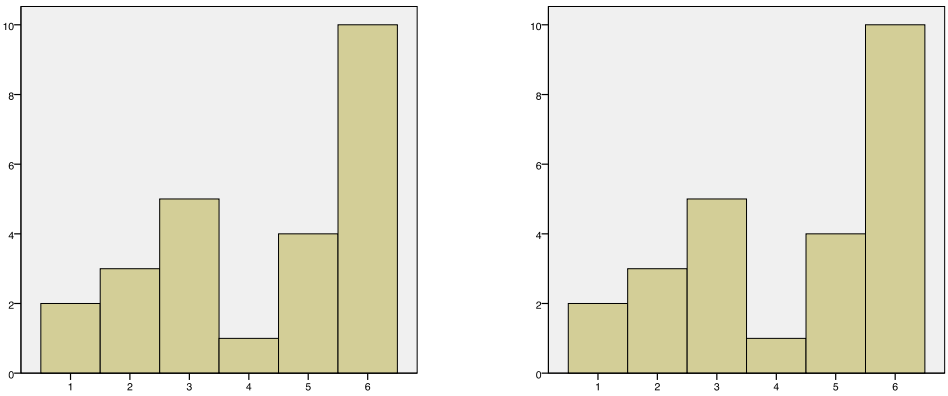

Figure 10. Score distribution for Savage for the Split Normal group(N=25) (left) and Full Normal group (N=49) (right)

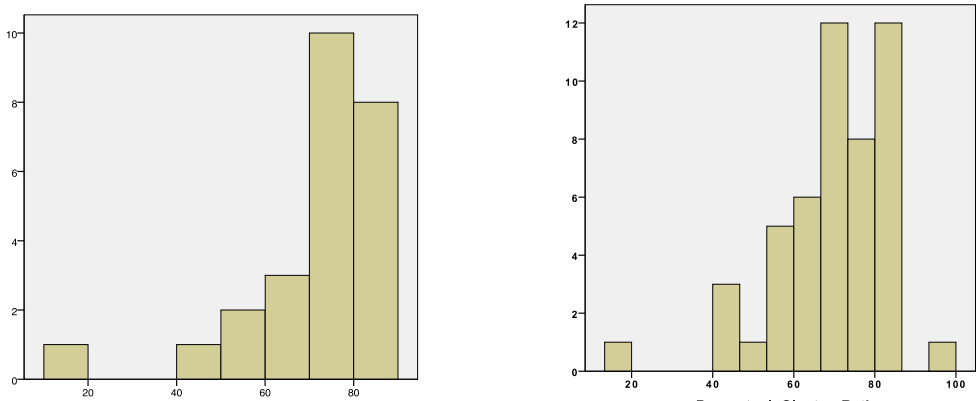

Figure 11. Score distribution for Shorr for the Split Normal group (N=25) (left) and Full Normal group (N=49) (right)

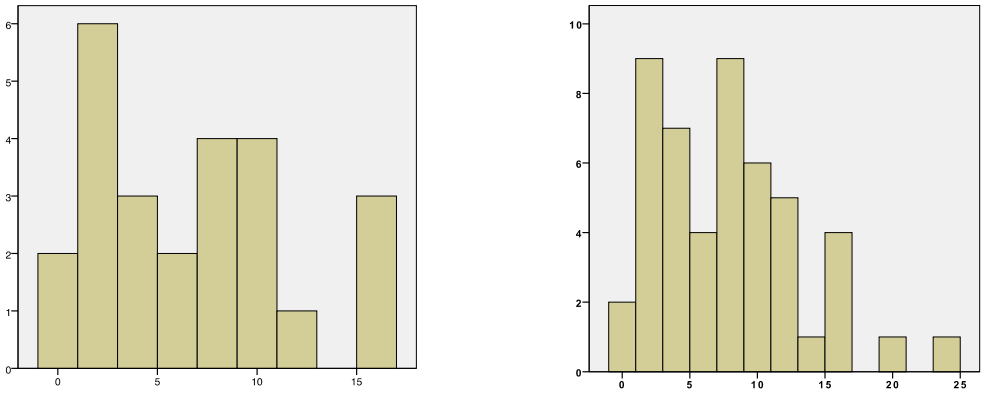

Figure 12. Score distribution for Visser for the Split Normal group (N=25) (left) and Full Normal group (N=49) (right)

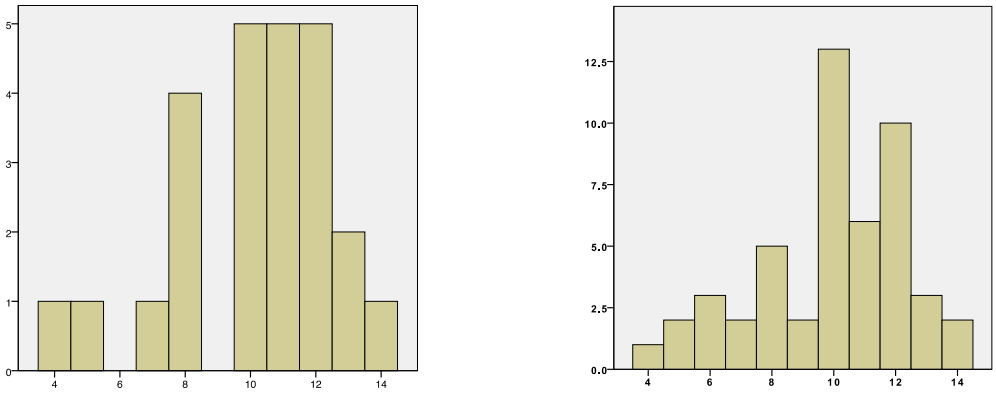

Figure 13. Score distribution for Waber for the Split Normal group (N=25) (left) and Full Normal group (N=49) (right)

Across the eight existing qualitative scoring systems examined, the distribution of scores obtained for the Full Normal group was largely commensurate with the findings reported by Troyer and Wishart (1997). Bennett-Levy's Total Strategy Scores and Bylsma's Q-Scores were normally distributed for the Full Normal group. This was also true for the Clinical group across both scoring systems, where there was also a wide distribution of scores. Bennett-Levy Total Strategy scores for the Wilson Normal group were, however, negatively skewed, revealing the strong influence of age on performance of the ROCF.

Osterrieth's Reproduction Types generated scores that were positively skewed for the Full Normal group and the Wilson Normal group, with most scores clustered around Types I and II. This was also the case for the Clinical group.

Savage et al's scoring system produced a greater range of scores, though still with a trend towards scores concentrated at high values for the Full Normal group and the Wilson Normal group. Participants in the Clinical group also

demonstrated greater variability of scores, though with less concentration of values at ceiling levels, as might be anticipated.

The distribution of scores for Shorr's Perceptual Cluster Ratio for the Full Normal group and the Wilson Normal group was relatively concentrated, with low score variability. This was less apparent for the Clinical group where a greater relative distribution of scores was evident, though the distribution of scores still deviated from normal. Visser scores were not transformed and therefore score distributions were positively skewed across all study groups. Waber and Holmes' Style scores were relatively negatively skewed across all study participants.

Scores for the novel process indices BAA and BAA-W were normally distributed for the Split Normal group and the Clinical group. A similar pattern was observed for the novel process indices, BAC and BAC-W. This result is noteworthy given the small sample size used for these descriptive analyses for the Split Normal group, and the comparability of findings for the larger Full Normal group.

Trends evident across visual inspection of histograms of score distributions were further reinforced by skewness and kurtosis statistics (Table 4) where the strengths of aforementioned existing qualitative scoring systems and quantified process measures was highlighted.

Table 4. *Results for Tests of Normality Across all Qualitative Scoring Systems*

Measure	Group	Skewness	Kurtosis	Kolmogorov-Smirnov		Shapiro-Wilk	
				Statistic	Sig.	Statistic	Sig.
BAC	Split Normal	0.211	-0.186	0.127	0.200	0.960	0.407
	Clinical	0.813	0.798	0.108	0.186	0.941	0.012
BAC-W	Split Normal	-0.191	-0.718	0.100	0.200	0.971	0.682
	Clinical	0.796	0.355	0.106	0.200	0.945	0.019
BAA	Split Normal	-0.390	-0.747	0.176	0.043	0.928	0.077
	Clinical	0.072	-0.713	0.087	0.200	0.979	0.485
BAA-W	Split Normal	-1.007	1.278	0.119	0.200	0.933	0.102
	Clinical	-0.151	-0.888	0.084	0.200	0.970	0.218
Bennett-Levy	Full Normal	-0.512	-0.414	0.143	0.014	0.960	0.093
	Split Normal	-0.763	0.501	0.146	0.178	0.948	0.222
	Clinical	-0.510	-0.557	0.115	0.082	0.960	0.078
	Wilson Normal	-0.689	0.092	0.116	0.083	0.953	0.042
Binder	Full Normal	-0.781	-0.675	0.294	0.000	0.795	0.000
	Split Normal	-0.837	-0.528	0.314	0.000	0.782	0.000
	Clinical	-0.397	-0.638	0.171	0.001	0.922	0.002
	Wilson Normal	-0.578	-0.623	0.222	0.000	0.855	0.000
Bylsma	Full Normal	0.187	-1.278	0.192	0.000	0.917	0.000
	Split Normal	0.070	-1.484	0.184	0.029	0.905	0.023
	Clinical	0.297	-0.766	0.104	0.200	0.968	0.168
	Wilson Normal	0.328	-0.911	0.132	0.027	0.946	0.022
Osterrrieth	Full Normal	1.724	2.666	0.325	0.000	0.662	0.000
	Split Normal	1.865	3.378	0.340	0.000	0.652	0.000
	Clinical	1.020	0.316	0.392	0.000	0.717	0.000
	Wilson Normal	1.483	1.769	0.274	0.000	0.699	0.000
Savage et al.	Full Normal	-0.369	-1.288	0.203	0.000	0.863	0.000
	Split Normal	-0.506	-1.261	0.231	0.001	0.832	0.001
	Clinical	0.058	-0.865	0.151	0.005	0.937	0.009
	Wilson Normal	-0.489	-0.761	0.196	0.000	0.902	0.000
Shorr et al.	Full Normal	-1.349	3.209	0.201	0.000	0.900	0.001
	Split Normal	-2.093	6.073	0.239	0.001	0.805	0.000
	Clinical	-0.573	0.179	0.142	0.010	0.956	0.052
	Wilson Normal	-1.026	1.706	0.164	0.002	0.932	0.006
Visser	Full Normal	0.830	0.488	0.110	0.183	0.936	0.011
	Split Normal	0.570	-0.536	0.141	0.200	0.924	0.064
	Clinical	0.737	-0.104	0.153	0.004	0.938	0.010
	Wilson Normal	0.523	-0.706	0.163	0.002	0.937	0.009
Waber &	Full Normal	-0.633	-0.137	0.204	0.000	0.939	0.014
Holmes	Split Normal	-0.885	0.557	0.201	0.011	0.925	0.067
	Clinical	-0.556	-0.457	0.131	0.027	0.949	0.027
	Wilson Normal	-1.052	2.228	0.148	0.007	0.921	0.002

Histograms are presented for each of the scoring systems for normal and clinical participants (Figures 14 -29), further reinforcing the pattern of score distributions.

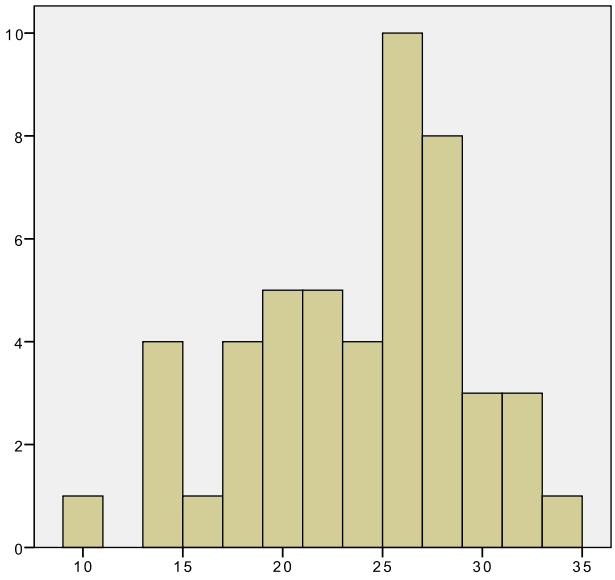

Figure 14. Frequency histogram of Bennett-Levy Strategy Scores for participants in the Full Normal group (N=49). Vertical axis shows frequencies out of 10 and horizontal axis shows

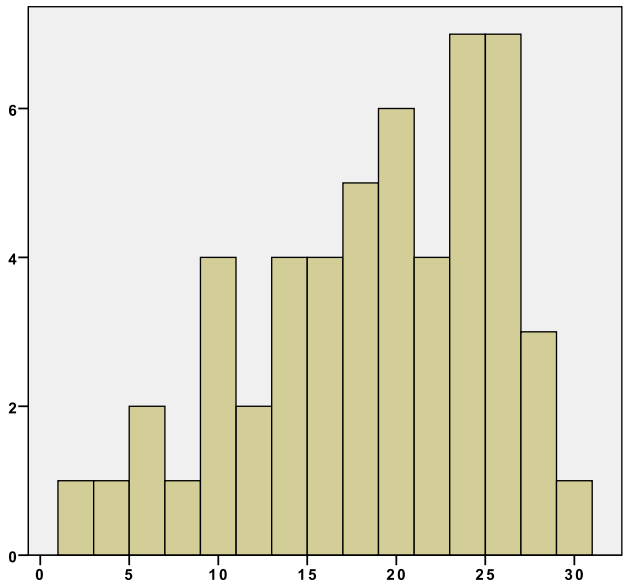

Figure 15. Frequency histogram of Bennett-Levy Strategy Scores for participants in the Clinical group. Vertical axis shows frequencies out of 8 and horizontal axis shows scores. scores.

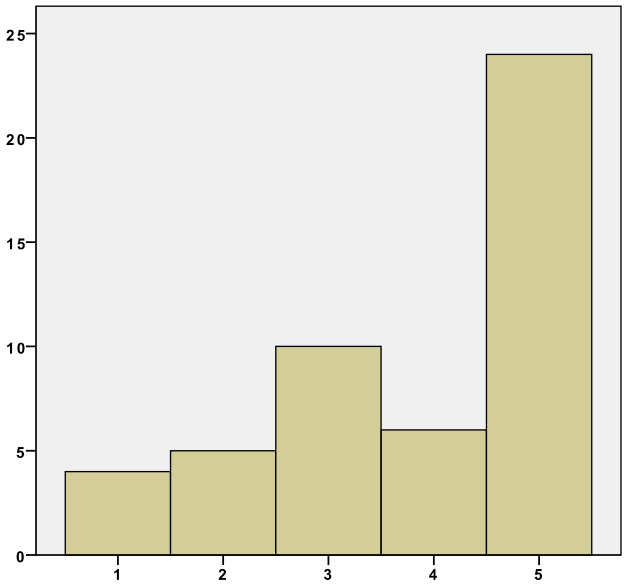

Figure 16. Frequency histogram of Binder Configural Scores for participants in the Full Normal group (N=49). Vertical axis shows frequencies out of 25 and horizontal axis shows

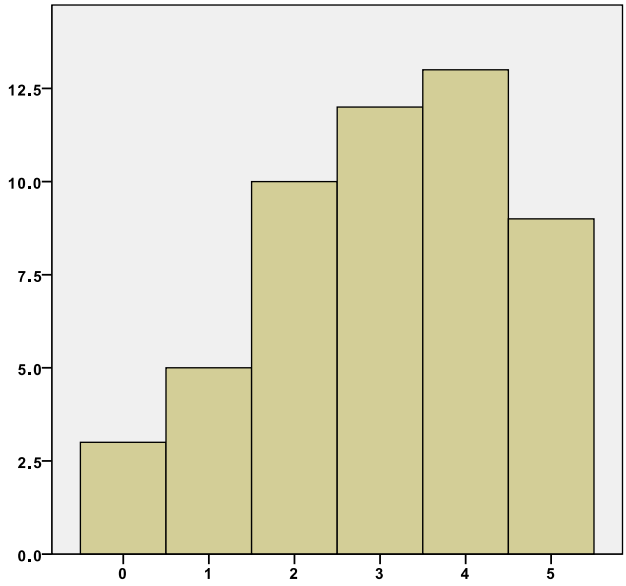

Figure 17. Frequency histogram of Binder Configural Scores for participants in the Clinical group. Vertical axis shows frequencies out of 15 and horizontal axis shows scores.

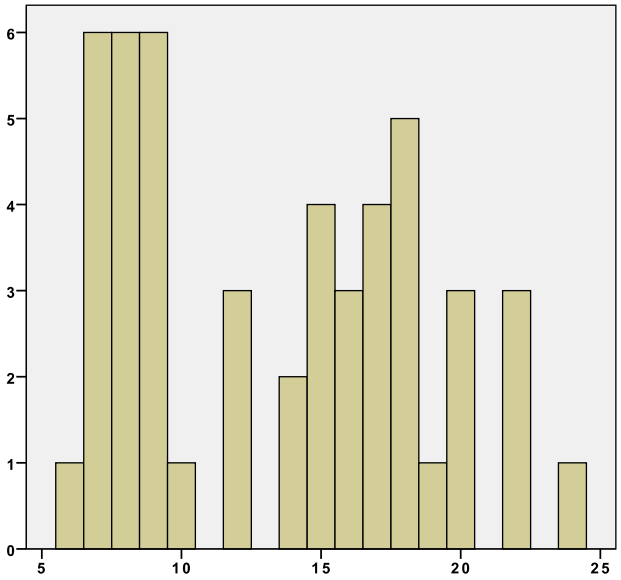

Figure 18. Frequency histogram of Bylsma Q-Scores for participants in the Full Normal group (N=49). Vertical axis shows frequencies out of 6 and horizontal axis shows scores.

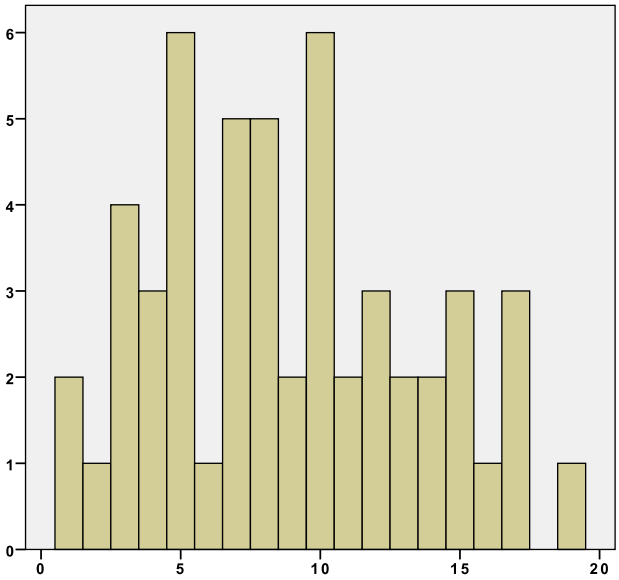

Figure 19. Frequency histogram of Bylsma Q-Scores for participants in the Clinical group. Vertical axis shows frequencies out of 6 and horizontal axis shows scores.

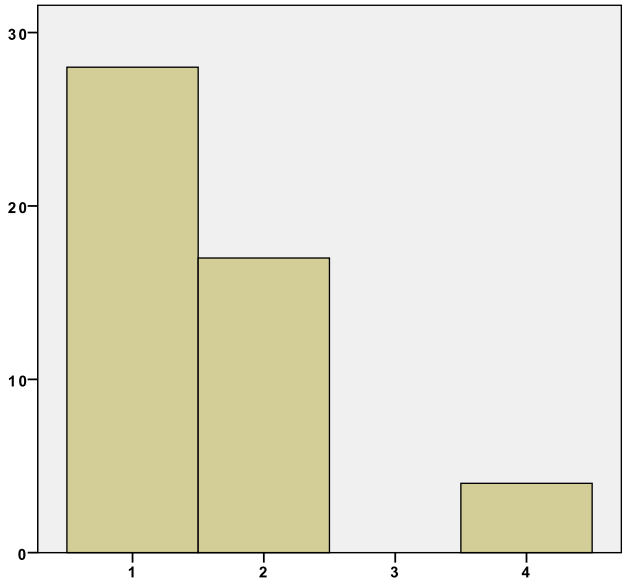

Figure 20. Frequency histogram of Osterrieth's Reproduction Types for participants in the Full Normal group (N=49). Vertical axis shows frequencies out of 30 and horizontal axis shows scores.

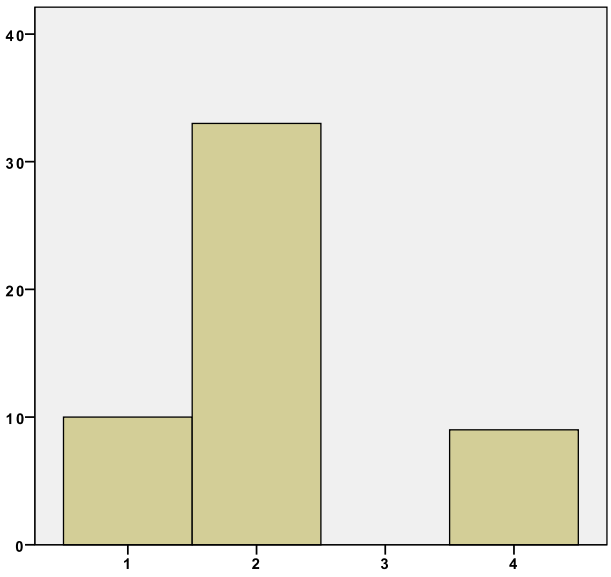

Figure 21. Frequency histogram of Osterrieth's Reproduction Types for participants in the Clinical group. Vertical axis shows frequencies out of 40 and horizontal axis shows scores.

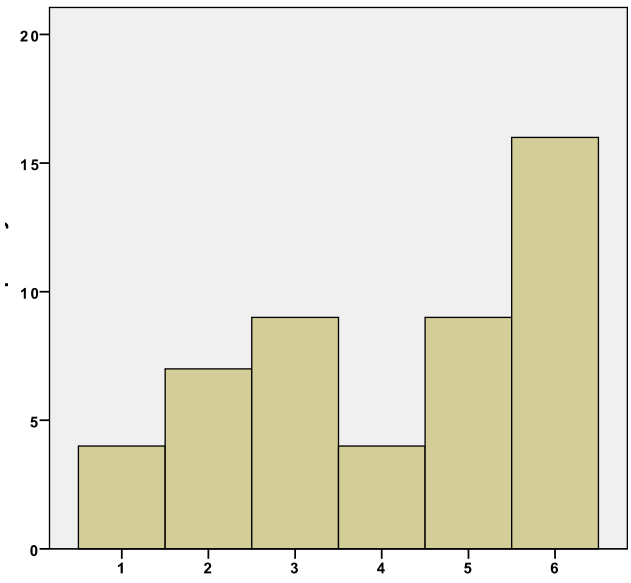

Figure 22. Frequency histogram of Savage Total scores for participants in the Full Normal group (N=49). Vertical axis shows frequencies out of 20 and horizontal axis shows scores.

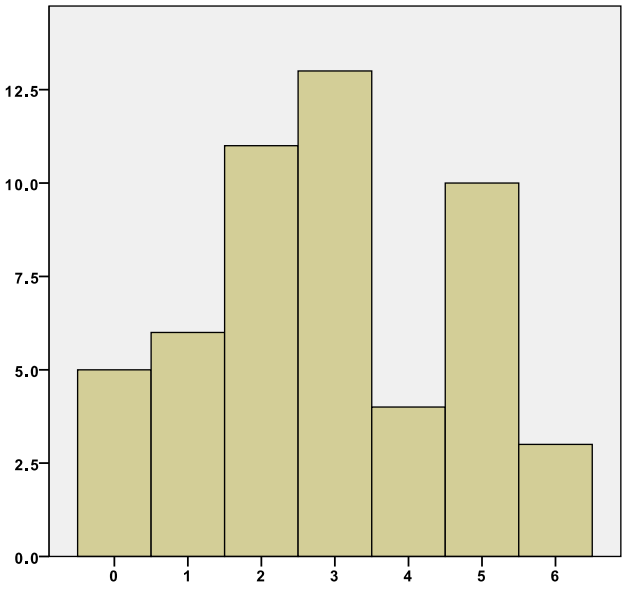

Figure 23. Frequency histogram of Savage Total scores for participants in the Clinical group. Vertical axis shows frequencies out of 15 and horizontal axis shows scores.

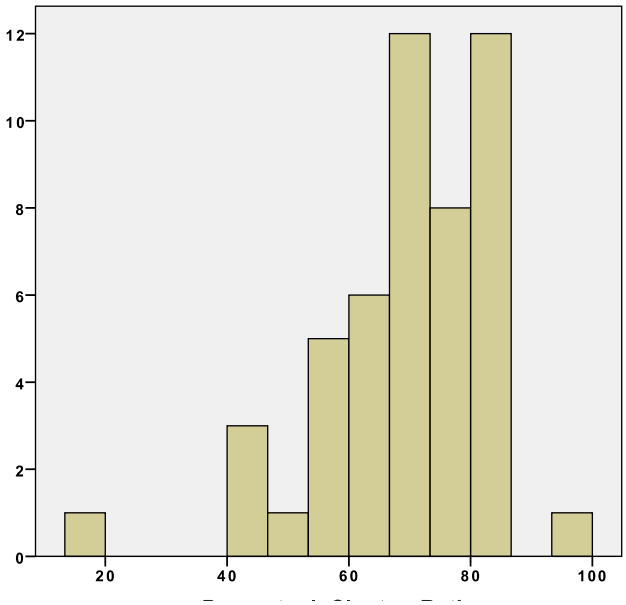

Figure 24. Frequency histogram of Shorr scores for participants in the Full Normal group (N=49). Vertical axis shows frequencies out of 12 and horizontal axis shows scores.

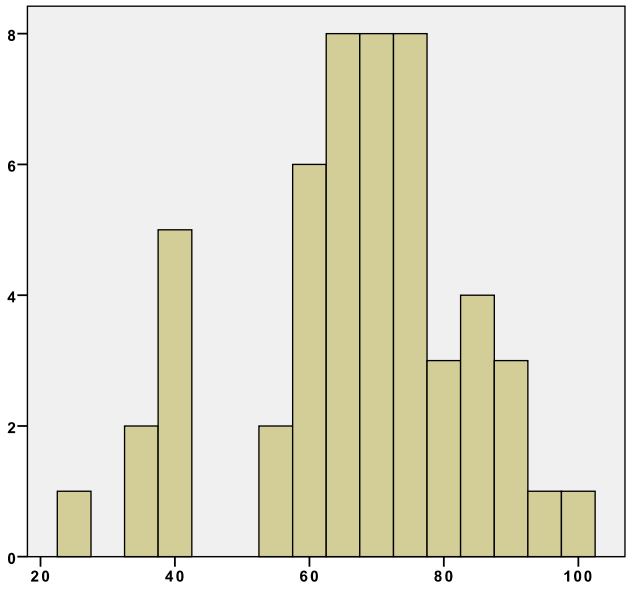

Figure 25. Frequency histogram of Shorr scores for participants in the Clinical group. Vertical axis shows frequencies out of 12 and horizontal axis shows scores.

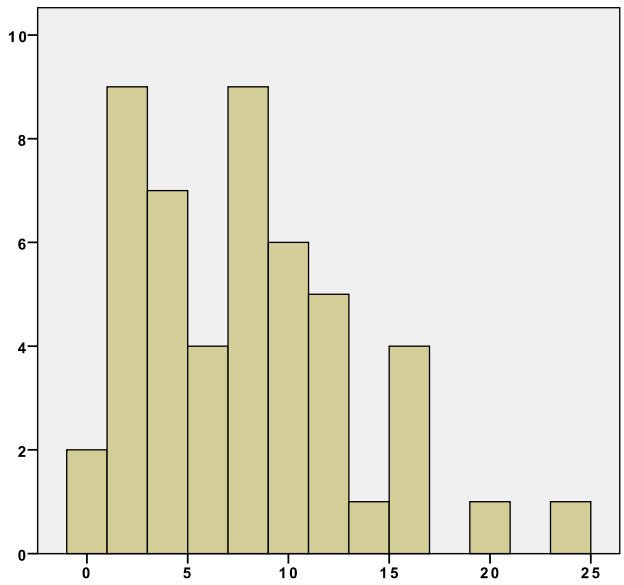

Figure 26. Frequency histogram of Visser scores for participants in the Full Normal group (N=49). Vertical axis shows frequencies out of 10 and horizontal axis shows scores.

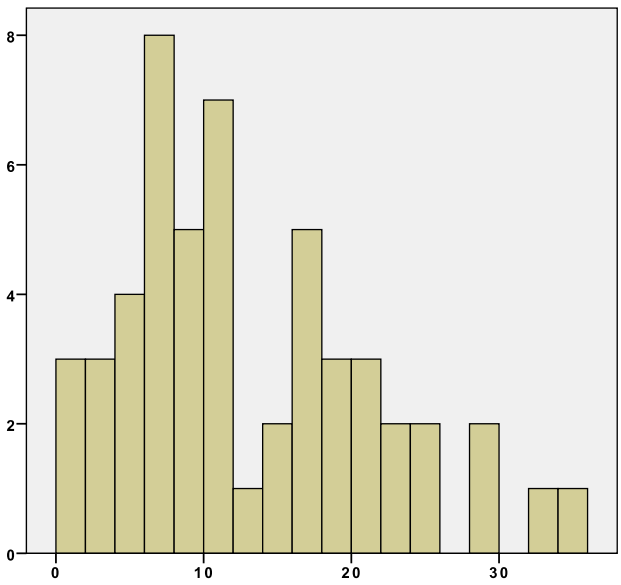

Figure 27. Frequency histogram of Visser scores for participants in the Clinical group. Vertical axis shows frequencies out of 8 and horizontal axis shows scores.

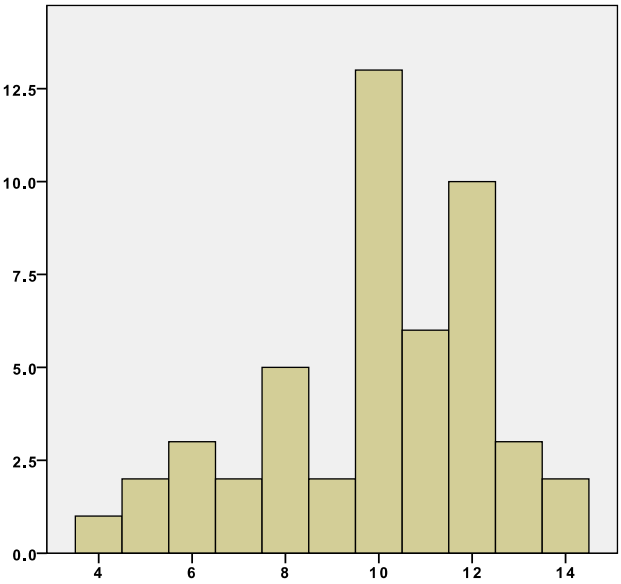

Figure 28. Frequency histogram of Waber Style scores for participants in the Full Normal group (N=49). Vertical axis shows frequencies out of 15 and horizontal axis shows scores.

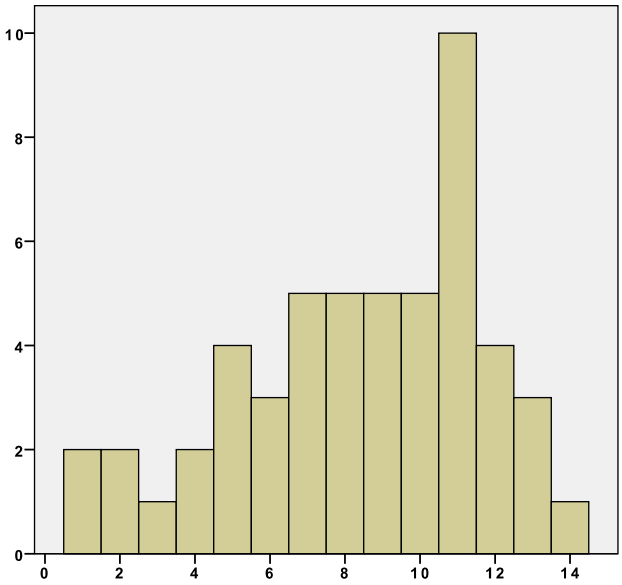

Figure 29. Frequency histogram of Waber Style scores for participants in the Clinical group. Vertical axis shows frequencies out of 10 and horizontal axis shows scores.

Scores for the novel process index, BAC were normally distributed across participants in the Split Normal group and the Clinical group as shown in Figure 30 and Figure 31.

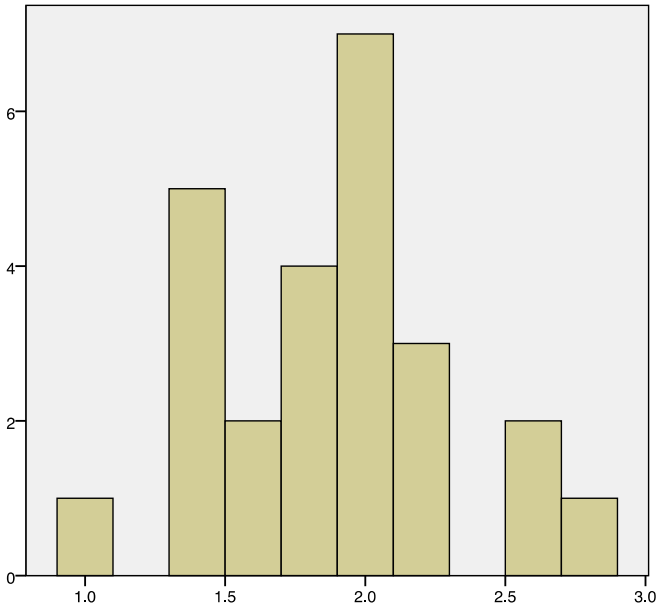

Figure 30. Frequency histogram of BAC scores for participants in the Split Normal group 1 (N=25). Vertical axis shows frequencies out of 8 and horizontal axis shows scores.

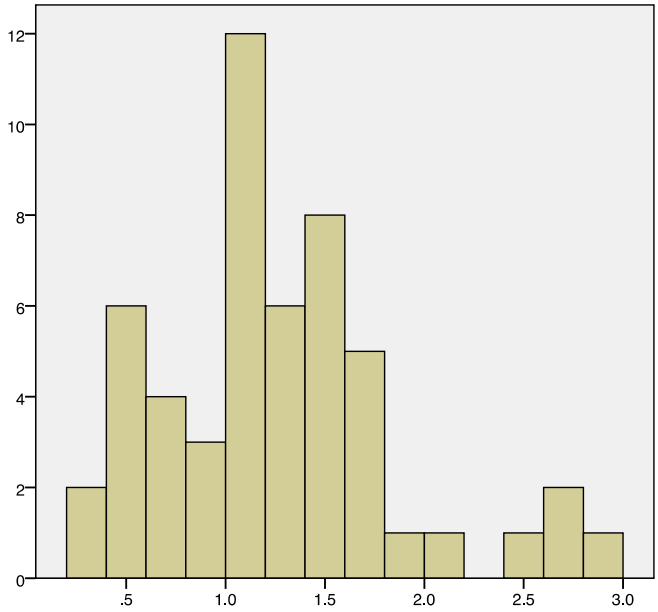

Figure 31. Frequency histogram of BAC scores for participants in the Clinical group. Vertical axis shows frequencies out of 12 and horizontal axis shows scores.

Scores for BAC-W were normally distributed across participants in the Split Normal group and the Clinical group as shown in Figure 32 and Figure 33, though with evidence of skewing for clinical participants.

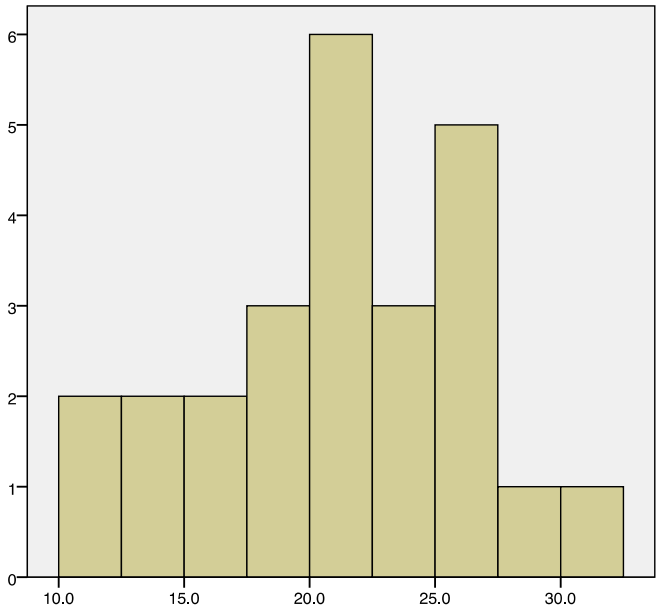

Figure 32. Frequency histogram of BAC-W scores for participants in the Split Normal group (N=25). Vertical axis shows frequencies out of 6 and horizontal axis shows scores.

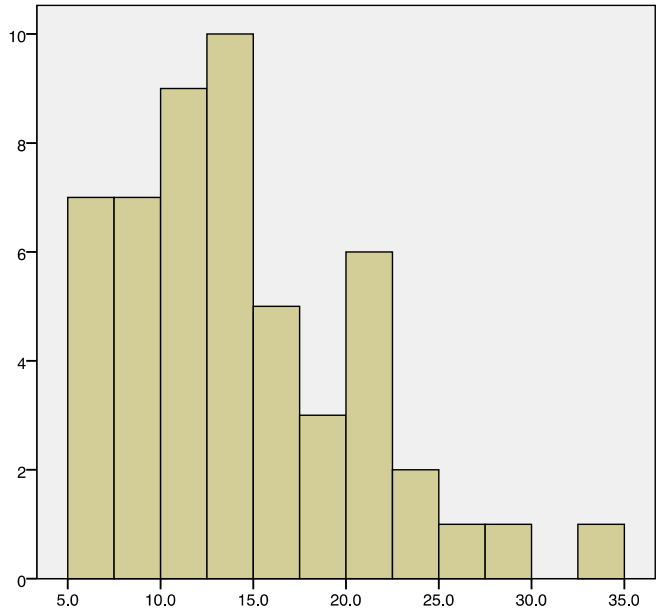

Figure 33. Frequency histogram of BAC-W scores for participants in the Clinical group. Vertical axis shows frequencies out of 10 and horizontal axis shows scores.

Score distributions for the novel process index, BAA were not normally distributed for participants in the Split Normal group (p=0.043) (Figure 34), while scores for the Clinical group were normally distributed as shown in Figure 35.

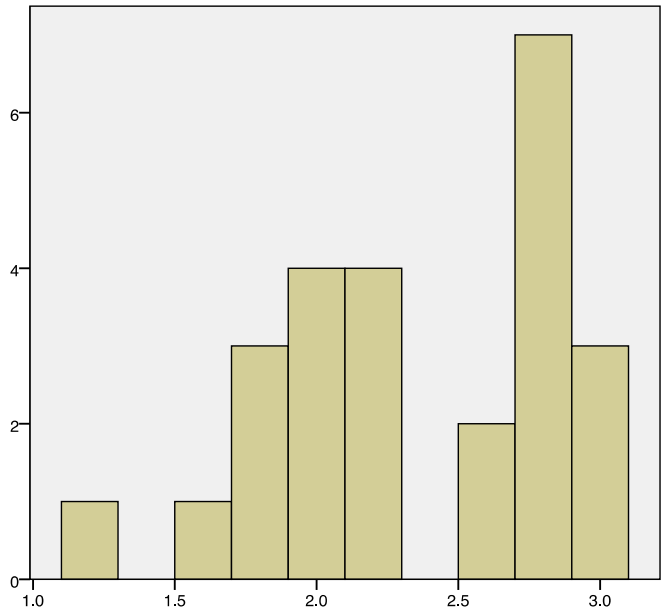

Figure 34. Frequency histogram of BAA scores for participants in the Split Normal group(N=25). Vertical axis shows frequencies out of 8 and horizontal axis shows scores.

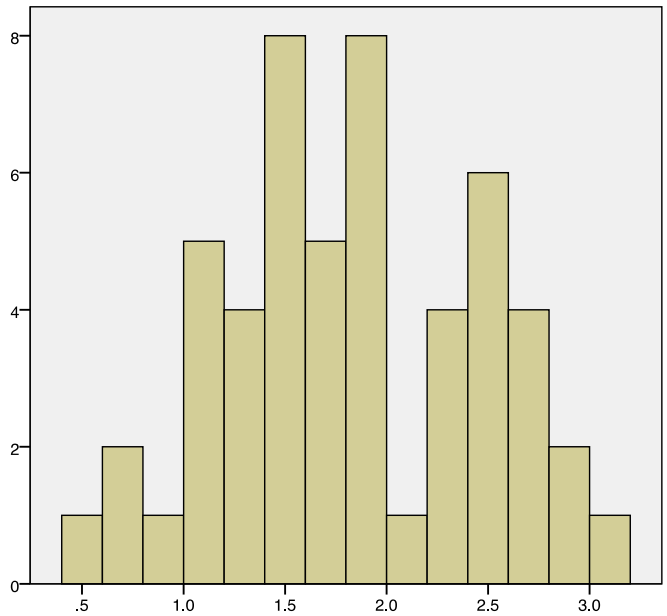

Figure 35. Frequency histogram of BAA scores for participants in the Clinical group. Vertical axis shows frequencies out of 8 and horizontal axis shows scores.

Score distributions for the weighted version of the novel process index BAA, BAA-W were normally distributed for participants across both the Split Normal and Clinical groups as shown in Figure 36 and Figure 37.

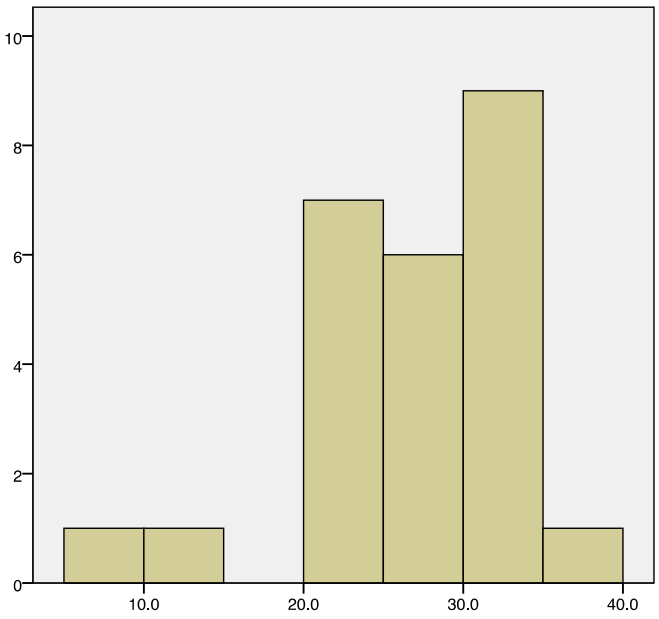

Figure 36. Frequency histogram of BAA-W scores for participants in the Split Normal group (N=25). Vertical axis shows frequencies out of 10and horizontal axis shows scores.

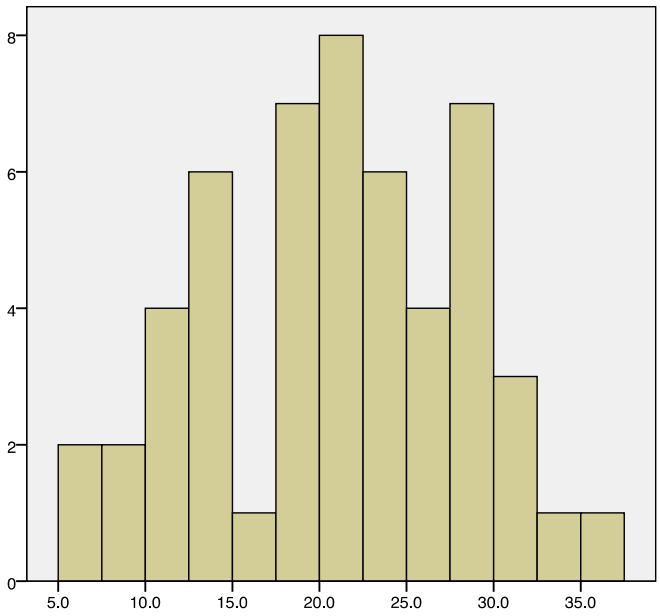

Figure 37. Frequency histogram of BAA-W scores for participants in the Clinical group. Vertical axis shows frequencies out of 8 and horizontal axis shows scores.

Descriptive analyses were undertaken on the Wilson group of Normal participants. A similar pattern of findings relative to the Full Normal group was obtained (Appendix 3)

A summary of qualitative scores obtained across study participants is provided in Table 5.

Table 5. *Qualitative Scores for all Study Groups*

Measure	Group	Mean	SD	Statistic	Sig. (2-tailed)
BAC	Split Normal	1.85	0.45	t(75) = 4.69	0.000
	Clinical	1.23	0.59		
BAC-W	Split Normal	21.0	5.48	t(75) = 4.61	0.000
	Clinical	14.3	6.26		
BAA	Split Normal	2.32	0.50	t(75) = 3.75	0.000
	Clinical	1.78	0.63		
BAA-W	Split Normal	26.4	6.23	t(75) = 3.18	0.002
	Clinical	20.8	6.96		
Bennett-Levy	Full Normal	23.1	5.53	t(99) = 3.98	0.000
	Clinical	18.2	6.90		
	Wilson Normal	22.9	4.92		
Binder	Full Normal	3.84	1.36	t(99) = 2.86	0.005
	Clinical	3.04	1.44		
	Wilson Normal	3.73	1.23		
Bylsma	Full Normal	13.5	5.19	t(99) = 4.70	0.000
	Clinical	8.85	4.64		
	Wilson Normal	13.5	4.86		
Osterrrieth	Full Normal	1.59	0.86	t(99) = -3.13	0.002
	Clinical	2.15	0.94		
	Wilson Normal	1.71	0.90		
Savage	Full Normal	4.12	1.74	t(99) = 3.54	0.001
	Clinical	2.90	1.72		
	Wilson Normal	4.18	1.52		
Shorr	Full Normal	68.3	14.1	t(99) = 4.42	0.679
	Clinical	67.0	16.5		
	Wilson Normal	60.3	13.7		
Visser	Full Normal	7.41	5.22	t(99) = -3.71	0.000
	Clinical	12.6	8.43		
	Wilson Normal	6.31	3.83		
Waber	Full Normal	9.94	2.41	t(99) = 2.68	0.009
	Clinical	8.38	3.32		
	Wilson Normal	9.31	2.10		

T-tests were only undertaken for the Split Normal group and the Clinical Group for the novel process measures, and for the Full Normal group and the Clinical group for all remaining scoring measures.

The mean scores obtained by participants in the Full Normal group were largely consistent with the findings reported by Troyer and Wishart (1997), with the exception of Shorr et al's Perceptual Cluster Ratio which was relatively lower, as was the Style Score for Waber & Holmes (1985, 1986) qualitative scoring system. Although participants in the Full Normal group in the current study were older than Bennet-Levy's normal sample (Mean age 29.3 years, SD: 9.3 years) the mean Total Strategy score of 23.1 (SD: 5.53) was comparable to the value reported in Bennett-Levy's sample (Mean = 23.4, SD: 5.0). Similarly, the mean organisation score reported for Savage's (1999) normal sample (Mean age 31.9 years, SD: 8.7 years) was 4.20 (SD: 1.51), comparable to current study findings. Statistically significant differences between participants in the Full Normal group and the Clinical group were evident across all scoring systems, with the exception of Shorr et al's scoring system.

Test Re-Test Reliability

It was hypothesised that quantified process measures would demonstrate at least moderate test-retest reliability. Study findings were not entirely supportive of this, as demonstrated in Table 6. Given the aforementioned pattern of differences in score distributions across each of the qualitative scoring systems under study, it was decided to include both Pearson and Spearman correlations.

Table 6. *Results for Test Re-Test Reliability Across all Scoring Systems*

Qualitative Scoring System	Pearson Correlation	Spearman Correlation
BAC	0.628	0.349
BAC-W	0.477	0.252
BAA	0.306	0.630
BAA-W	0.078	0.180
Bennett-Levy	0.779*	0.758*
Binder	0.755*	0.618
Osterrieth	0.104	0.370
Savage	0.660	0.566
Bylsma	0.716*	0.612
Shorr et al.	0.773*	0.632
Visser	0.776*	0.521
Waber & Holmes	0.926**	0.891**

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

The results for re-test reliability revealed significant results for each of the existing qualitative scoring systems, with the exception of Osterrieth's Reproduction Types, which performed poorly on re-test and Savage et al's (1999) scoring system ($r=0.66$, $p=0.075$). Waber & Holmes' (1985) Style Score demonstrated the strongest stability ($r=0.926$, $p<0.01$). The quantified process measures did not perform strongly on re-test relative to existing scoring systems, contrary to expectations. BAC was identified as the strongest index in this context, ($r=0.628$, $p=0.096$), while BAA-W performed most poorly in comparison to other quantified process measures and all other qualitative scoring systems reviewed ($r=0.078$, $p=0.854$).

Sensitivity

It was hypothesised that quantified process measures would discriminate between participants in the Split Normal group and participants in the

Clinical group. Logistic regression analyses were undertaken to examine whether scores for the existing qualitative scoring systems and the quantified process measures could predict clinical status. A summary of the data is presented in Table 7.

Table 7. *Summary of Logistic Regression Analyses of Clinical Group Status as a Function of Scores Obtained on Qualitative Measures*

Predictors	B	SE	Wald	Odds Ratio	Sig. (2-tailed)	95% CI for Odds Ratio Lower Bound	Upper Bound
BAC	-2.016	0.548	13.558	0.133	0.000	0.046	0.389
BAC-W	-0.174	0.047	13.771	0.840	0.000	0.767	0.921
BAA	-1.577	0.484	10.618	0.207	0.001	0.080	0.533
BAA-W	-0.118	0.042	8.131	0.888	0.004	0.819	0.964
Bennett-Levy	-0.172	0.053	10.773	0.842	0.001	0.759	0.933
Binder	-0.450	0.195	5.305	0.638	0.021	0.435	0.935
Savage	-0.453	0.154	8.642	0.636	0.003	0.470	0.860
Bylsma	-0.200	0.055	13.46	0.819	0.000	0.736	0.911
Osterrieth	0.886	0.369	5.762	2.426	0.016	1.177	5.002
Shorr	-0.009	0.016	0.335	0.991	0.563	0.961	1.022
Visser	0.140	0.047	8.748	1.151	0.003	1.048	1.263
Waber & Holmes	-0.205	0.094	4.795	0.814	0.029	0.678	0.979

As detailed in Table 7, all of the qualitative scoring systems and quantified process measures made a statistically significant contribution to the prediction of clinical status, with the exception of Shorr et al.'s (1992) scoring system. As expected, a negative relationship between low scores and clinical status was revealed, with the exception of Visser's (1973) scoring system and Osterrieth's (1944) Reproduction Types where elevated scores reflected performance deficits. Results for classification accuracy are presented in Table 8

Table 8. *Summary of Classification Rates for all Qualitative Scoring Systems*

Scoring System	Percentage Correct		Overall Percentage
	Split Normal Group (N=25)	Clinical Group (N=52)	
BAC	52	88	76.6
BAC-W	40	88.5	72.7
BAA	44	86.5	72.7
BAA-W	36	90.4	72.7
Bennett-Levy	52	86.5	75.3
Binder	0	100	67.5
Savage	40	94.2	76.6
Bylsma	48	90.4	76.6
Osterrieth	0	100	67.5
Shorr	0	100	67.5
Visser	36	88.5	71.4
Waber & Holmes	12	92.3	66.2

Logistic regression analyses demonstrated that the overall classification rate for the quantified process measures was comparable to the stronger existing qualitative scoring systems, those of Savage, Bylsma and Bennett-Levy. The index BAC featured strongly in this regard, demonstrating an overall classification percentage of 76.6, which was equivalent to the results for the Savage and Bylsma scoring approaches. The other quantified scoring measures all demonstrated a classification percentage of 72.7, which surpassed the results of many of the existing scoring systems under review. The weakest scoring systems as identified by these analyses were those of Binder, Shorr, and Waber and Holmes, with specificity being especially poor for those scoring approaches.

Relationship Between Planning and Recall

Quantified process measures were posited to demonstrate an association with recall, with at least moderate correlations observed across control and clinical participants. Correlations were calculated between qualitative scoring system total scores and recall scores. Results are presented in Table 9. Again, both Pearson and Spearman correlations were provided for each scoring index.

Table 9. *Correlation Matrix for Qualitative Scoring Systems and Recall*

Measure	Group	Pearson	Spearman
BAC	Split Normal	0.476*	0.541**
	Clinical	0.445**	0.448**
BAC-W	Split Normal	0.569**	0.620**
	Clinical	0.351*	0.359
BAA	Split Normal	-0.391	-0.409*
	Clinical	0.454**	0.467**
BAA-W	Split Normal	-0.289	-0.414*
	Clinical	0.425	0.432**
Bennett-Levy	Full Normal	0.787**	0.802**
	Clinical	0.467**	0.444**
Binder	Full Normal	0.662**	0.698**
	Clinical	0.387**	0.367*
Bylsma	Full Normal	0.778**	0.733**
	Clinical	0.518**	0.537**
Osterrrieth	Full Normal	-0.556**	-0.742**
	Clinical	-0.162	-0.143
Savage et al.	Full Normal	0.740**	0.737**
	Clinical	0.429**	0.419**
Shorr et al.	Full Normal	0.436*	0.425*
	Clinical	0.149	0.183
Visser	Full Normal	-0.648**	-0.624
	Clinical	-0.573**	-0.602
Waber-Holmes	Full Normal	0.554**	0.578**
	Clinical	0.455**	0.437**
Accuracy	Full Normal	0.541**	0.525**
	Clinical	0.716**	0.726**

**Correlation significant at the 0.01 level (2-tailed)

* Correlation significant at the 0.05 level (2-tailed)

Analyses revealed significant, moderate to strong correlations between existing qualitative scoring systems and recall, in keeping with the hypothesised relationship between copy strategy and recall. Also within expectation was the negative relationship between higher scores on Visser's scoring system and Osterrieth's Reproduction Types, where higher values denoted poorer copying strategies. Of the quantified process measures developed, the strongest correlations were revealed for BAC across participants in the Clinical group ($r=0.445$, $p<0.01$) and for participants in the Split Normal group ($r=0.476$, $p<0.01$). Significant correlations were also revealed for BAC-W across participants in the Split Normal group ($r=0.569$, $p<0.01$) and participants in the Clinical group ($r=0.351$, $p<0.05$). In contrast, a negative correlation was demonstrated between BAA and recall for participants in the Split Normal group, and similarly for BAA-W for participants in the Split Normal group. This stands in contrast to the significant positive relationship between BAA and recall for clinical participants ($r=0.454$, $p<0.01$), and the positive relationship between recall and BAA-W for clinical participants, which was not, however, significant. Again, of the existing qualitative scoring systems examined, Bennett-Levy's Total Strategy Score demonstrated the strongest correlation with recall, with Bylsma's Q-score and Savage et al's (1999) Organisation score also demonstrating strong correlations with recall. Analysis of the results overall reveals that all scoring systems were at least moderately related to recall, with the exception of Osterrieth's Reproduction Types for Clinical participants.

Discussion

An earlier comprehensive investigation of the psychometric properties of existing qualitative scoring systems for the ROCF had identified that some of these scoring systems were superior to others in relation to distribution of scores generated and the association between identified process variables and recall (Troyer & Wishart, 1997). The current study identified findings that largely parallel this, even though the normal sample comprised younger participants than Troyer and Wishart's (1997) study. Despite this age difference, score distributions similar to Troyer and Wishart's normal cohort were revealed for participants in the Full Normal group, (Age Range: 18-82 years) and for the Wilson Normal group of undergraduate students (Age Range: 18-34 years). Further, the descriptive analyses for the Wilson Normal group also revealed a similar trend towards superior score distributions for Bennett-Levy Strategy Scores and Bylsma Q-Scores. Not surprisingly, the categorical rating systems (Osterrieth's Reproduction Types and Binder's Configural Score), were demonstrated to produce skewed distributions, with scores concentrated at high values. Of note, Savage et al's scoring system, which is largely based on Binder's scoring approach but with the addition of the central rectangle, produced a greater range of scores relative to Binder's Configural Score, though scores were also not normally distributed. Scoring distributions for Shorr et al's Perceptual Cluster Ratio and Waber and Holmes' Style Score were clustered, with low score variability for both groups of normal participants. One difference which was observed was the greater range of scores for Visser's scoring system for participants in the Wilson Normal group relative to older participants in the Full Normal group, though the score distributions deviated from normal for both cohorts.

The hypothesised normal distribution of scores for the quantified process measures was confirmed in the current study. Scores derived for participants in the Split Normal group were normally distributed across all four quantified process measures, with a good range of scores demonstrated. This is encouraging and suggests the efficacy of empirically derived measures in producing score distributions which are amenable to statistical analysis.

Development and validation of quantified process measures under investigation required that the Full Normal group of 49 participants be divided into two smaller groups, resulting in a loss of statistical power for subsequent analyses. Importantly, the two split groups were matched for age and education. Reference to the distribution of scores for the Split Normal group (N=25) on which the quantified process measures were validated revealed similar score distributions relative to the Full Normal group (N=49) from which these groups were formed. Further, normal score distributions were demonstrated. This result was important in establishing the validity of this reference group for analyses subsequently undertaken as the loss of power represented a notable limitation.

A strong association has been demonstrated between age and performance on the ROCF (Bennett-Levy, 1984, Berry et al, 1991; Boone et al, 1993; Hartman & Potter, 1998; Mitrushina et al, 1990). The advantage offered by inclusion of the younger Wilson Normal group in the current study is that it provided insight into how score distributions for qualitative scoring systems under review might be influenced by age. Despite the significant difference in age between this group and

the Full Normal Group, score distributions were largely consistent across the majority of the qualitative scoring systems investigated. This suggests the relative stability of score profiles across each of the qualitative scoring systems examined. Further investigation of this trend is required with a much larger subject sample across a broader range of age categories. This would enable a more refined analysis of score variability, leading to further delineation of how scores across these qualitative scoring systems are influenced by age. A broader sample of normal participants would also enable an examination of the influence of other demographic variables such as gender and education, where research findings generally remain inconclusive.

Re-test reliability is a central construct in the clinical context of ROCF assessment. An important consideration when examining test-retest data for the ROCF is the effect of prior exposure and the loss of novelty. Reliability values for copy trials using the standard scoring system in normal populations have ranged from poor (r=0.18, Berry et al, 1991), to moderate (r=0.56-0.68, Mitrushina & Satz, 1991), to strong (r=0.94, Tupler et al, 1995). Bearing in mind the limitations of the small sample size used in the current study (N=8), reliability coefficients were not in the expected range for quantified processing measures, which stands in contrast to favourable findings demonstrated for other qualitative scoring systems under review. Of the qualitative scoring systems examined, Osterrieth's Reproduction Types demonstrated the lowest reliability in relation to retest changes on copy trial of the ROCF (r=0.370), while Savage et al's (1999) system demonstrated moderate reliability, without reaching statistical significance (r=0.66). The scoring approach demonstrating the best stability of scores on repeat administration of the ROCF was Waber & Holmes' Style score (r=0.926). All other existing qualitative scoring

approaches were also significant at the 0.05 level. While the quantified process measure BAC revealed a moderate correlation, (r=0.628) as did the weighted variant of this index (BAC-W: r=0.477), these measures did not reach statistical significance. The BAA index performed very poorly in terms of temporal stability (r=0.306), as was the case for the weighted version of this index, BAA-W (r=0.07), which demonstrated the lowest reliability coefficient across examination of all scoring systems. Given the extremely small sample size used for examining test-retest improvements in the current study, retest reliability for all scoring systems needs to be more comprehensively investigated in a much larger sample of healthy controls. Inclusion of a broader age range will also facilitate better analysis of whether practice effects vary as a function of age.

An important aspect of any qualitative scoring system for the ROCF is that the measures that are derived discriminate between normal controls and clinical participants. The results of the current study provided evidence in support of this across the qualitative scoring systems examined, and further confirmed the hypothesised ability of quantified process measures to contribute significantly to the prediction of clinical status. There were clear, significant differences between normal and clinical participants across all qualitative scoring systems examined, with the exception of Shorr et al's (1992) Perceptual Cluster Ratio, where statistically significant differences were not observed in the current study. Logistic regression analyses further revealed that the three least effective scoring scorings for correctly classifying clinical status were those of Waber and Holmes (1985, 1986) (Overall Classification Percentage: 66.2), Binder (1982) (Overall Classification Percentage: 67.5), and Shorr et al. (1992) (Overall Classification Percentage: 67.5). While these

systems could reliably classify clinical participants, specificity was extremely poor with all participants in the Split Normal Group incorrectly classified as clinical participants. These systems also featured poorly in analyses of score distributions for the clinical sample, where variability between participants was generally not demonstrated and scores were either clustered centrally or at extreme values. Binder's scoring system has been criticised because it fails to incorporate much of the important detail in the ROCF (Shorr et al, 1992; Hamby et al, 1993) and this might account in part for the pattern of findings observed. Most subjects achieved high values on copy trial, without apparent discrimination of the range of approaches to reproducing the ROCF. As already noted, Savage et al's (1999) scoring system, which differs only in the inclusion of the central rectangle performed more strongly in predicting clinical status, achieving an overall percentage classification of 76.6, equivalent only to Bylsma and the quantified process index BAC. This is not unexpected in light of Osterrieth's (1944) assertion that these elements form the "central armature" (p. 5) of the ROCF (cited in Corwin & Bylsma, 1995). The inclusion of additional detail within the configural structure of the ROCF did not, however, appear to enhance the predictive power of Shorr et al's (1992) Perceptual Cluster Ratio. The overall classification percentage for this system was only 67.5% and all of the participants in the Split Normal group were incorrectly classified as clinical participants.

It is perhaps not altogether surprising that Waber & Holmes' (1985, 1986) Style score did not feature strongly in the current study in relation to predicting clinical status (Overall Classification Percentage: 66.2) nor in analyses of score distribution/range, with evidence of peaked scores at maximal values. Similar

findings were reported in Troyer and Wishart's (1997) review where the Style score was also used in isolation to other measures from the broader scoring system from which this score was taken. Waber and Holmes (1986) recommended that the clinical use of the DSS-ROCF should not be restricted to organisation and style ratings and that other parameters, including assessment of accuracy and error patterns equally contribute to diagnostic formulations. It is also important to emphasise that the DSS-ROCF was developed for use in paediatric populations, limiting inferences that can be drawn from only one of a number of measures in this scoring system applied to adults. The strengths of the DSS-ROCF would appear to be better revealed in paediatric populations where the influence of developmental factors serve as the context for interpretation of copy approach to the ROCF.

The existing qualitative scoring systems identified to have the strongest classification accuracy were represented by those developed by Savage et al (1999), Bylsma et al, (1995) and Bennett-Levy (1984), each of these systems demonstrating an overall classification accuracy of 76.6. The latter two of these systems were also identified to be strong psychometrically in Troyer and Wishart's (1997) analysis. As hypothesised, the quantified process measure BAC was equally strong in relation to classification accuracy, also achieving an overall classification accuracy of 76.6%. The other quantified process measures also performed solidly in classifying clinical status, each demonstrating an overall classification accuracy of 72.7%. These findings further highlight the merits of these empirically derived quantified process measures in capturing copy strategy and planning on the ROCF. The advantage offered by these measures is that in larger numbers, across a wider

range of age bands, they potentially have the capacity to more closely approximate copy approach for a normal reference sample, thereby increasing sensitivity.

The central focus for analysis of recall performance in the current study related to how information from the ROCF is encoded and whether this relates to storage and retrieval. Given the complexity of the ROCF stimulus, adequate recall relies upon a framework where the simple elements of the design are accurately recalled relative to the spatial location of overarching elements. Study findings replicated Troyer and Wishart's (1997) analysis which identified that all of the reviewed qualitative scoring systems were moderately related to recall across Normal participants (Range: 0.30 - 0.47), though current study findings revealed somewhat higher correlations (Range: 0.436 – 0.787). Moderate correlations were also identified for clinical participants in the current study (Range: 0.149 – 0.518). The only exception was for Osterrieth's Reproduction Types which, although significantly correlated with recall for participants in the Full Normal group, was not significantly associated with recall for Clinical participants. Again, the highest correlations for participants in the Full Normal group were obtained using the qualitative scoring systems developed by Bennett-Levy (1984) (r=0.787), Bylsma et al. (1995) (r=0.778), and Savage et al. (1999) (r=0.740). Accuracy scores derived from the standard scoring system were significantly correlated with recall accuracy scores for participants in the Full Normal group (r=0.541) and participants in the Clinical group (r=0.716). Accuracy scores were not considered in correlation analyses in Troyer and Wishart's (1997) study. Among the quantified process measures, the weakest findings were again evident for BAA, which was only correlated with recall for participants in the Clinical group (r=0.454), but not for participants in the Split Normal group (r=-

0.391). It is difficult to account for the negative correlations between the measures BAA and BAA-W with recall for participants in the Split Normal group. The predicted relationship would have been a positive correlation with recall such that higher scores would represent performances more in keeping with the reference group from which these indices were derived. BAC demonstrated the strongest association with recall both for participants in the Split Normal group (r=0.476) and participants in the Clinical group (r=0.495). These results collectively highlight the importance of copy strategy for recall of the ROCF, and support the efforts made to better understand processes measures which contribute to encoding and thereby influence recall. It is encouraging that the quantified process measure BAC is comparable to other existing qualitative scoring measures in this context.

Binder's (1982) qualitative scoring system has made an invaluable contribution to process analysis for the ROCF. The configural elements of the ROCF thus defined have featured strongly in other scoring approaches such as Bylsma et al's (1995) Q-Score, Shorr et al's (1992) Perceptual Cluster Ratio, and of course Savage et al's (1999) scoring approach. While there is recognition that drawing the configural units of the ROCF in an unfragmented manner is important, more recent approaches have attempted to delineate other associated factors, which provide richer clinical information. Certainly, it would appear that drawing the configural elements at the beginning of the reproduction gives structure, which then facilitates better planning and execution of drawing approach. But in and of itself, this formulation of copy approach is not sufficient to capture other parameters which also contribute to variance in task performance such as fragmentation. Delis (1989) identified that one of the limitations of the ROCF is that there is no clear differentiation between

stimulus features perceived as “larger wholes versus smaller parts” (p16). Thus, while the central rectangle can be identified as a configural element and the features of the dots in the circle as ‘internal details’, it is difficult to unambiguously define other features contained within, such as the diagonals, as configural or detail. Delis argues that the ROCF does not “lend itself to rigorous quantification of performance differences in constructing configural versus detail features” (p16.) This might account for some of the difficulties inherent in qualitative scoring systems based solely on this premise.

It is important to replicate in clinical samples, findings from normal samples. The current study is the first to apply Bennett-Levy’s (1984) qualitative scoring system to a clinical population. In this context, Bennett-Levy’s system has demonstrated equal merit when applied to clinical groups as it has to normal populations, with a demonstrated ability to produce scores that vary between participants with a range of clinical presentations. One of the difficulties with this system is that while the Continuation Score was easy to derive, the Symmetry Score was more difficult to translate from information detailed in Bennett-Levy’s (1984) study and required significant time and deliberation to define scoring criteria relative to the other scoring systems which also performed strongly. It would be of considerable benefit if the Symmetry Score was to be better translated, thereby reducing the amount of subjectivity currently required to interpret this information. This would further enable clinicians and researchers to more readily use this measure as intended by the author in order to assess copy strategy.

Apart from Troyer and Wishart's (1997) analysis, there has been a dearth of research establishing the equivalence of scores obtained from Visser's scoring system and other qualitative scoring systems. Current findings demonstrate the comparability of scores derived from Visser's system to other scoring systems using the traditional administration of the complex figure in both normal and clinical populations. This is perhaps not surprising given that figure orientation has been reported to show minimal influence on copy performance of the ROCF (Ferraro et al, 2002). Although score distributions for the Visser system were skewed, Visser scores were significantly correlated with recall in normal and clinical participants, contributed significantly to the prediction of clinical status and demonstrated an Overall Classification Percentage of 71.4 %.

Past research examining the frequency of Osterrieth's Reproduction Types in normal samples had revealed smaller percentages of Reproduction Types I and II relative to the figure of 83% quoted in Osterrieth's (1944) original normative sample, cited in Lezak et al, (2012). In contrast to Maillet's (1992) finding that only 59% of adults in a sample aged 16-80 years (Mean age: 32.6 years) adopted Reproduction Types I and II, in the current sample, 92% of adults aged 18-82 years (Mean Age: 46.9 years) adopted Reproduction Types I and II. Only 8% of normal participants adopted Reproduction Types IV-VII, though this was a combined category due to the decision not to consider accuracy of copy approach, unlike other studies where Osterrieth's original seven Reproduction Types were preserved (Maillet, 1992, Ska & Nespoulos, 1988). Reproduction Type III was not observed for the Normal groups in the current study, in keeping with Osterrieth's (1944) observation cited in Lezak et al, (2012) of only one adult subject demonstrating

Reproduction Type III. It would be informative to examine the stability of frequency patterns of Osterrieth's Reproduction Types across a larger sample of normal participants.

One issue which arose indirectly in the context of examining flowchart representations of copy strategy for the ROCF, particularly for a subset of Clinical participants, was consideration as to whether impaired ability represented poor planning versus visuospatial dysfunction. The study investigator was blinded to the aetiology of acquired brain impairment for Clinical participants and therefore it was not possible or relevant to address this issue for the purposes of the current analysis. Future research examining a range of clinical groups will need to address the potential influence of other cognitive impairments on copy strategy for the ROCF. For some of the Clinical participants in the current study there were significant omissions with very few details of the ROCF reproduced on copy, let alone recall. In these cases, it would appear that the ROCF might not represent the best measure of planning and/or recall. If quantified process measures are to be used as measures of planning, then it will be important to appropriately target these measures for patients with acquired brain impairment, thus reducing the ambiguity in determinations of suboptimal performance.

The quantified process measures developed in the current study are intended to capture performance parameters relating to planning strategy rather than spatial organisation. Considerable attention was given to the concept of fragmentation. Thus, once construction of a particular element was commenced, regardless of whether it was one of the configural units or other elements of the

figure, it was imperative that construction was continued through to completion of this element. Indeed, many of the existing qualitative scoring systems under investigation place emphasis on the continuous/contiguous completion of designated elements, the only difference being that other aspects of the construction are also highlighted. Visser's position on evaluative systems for the ROCF was that the individual's approach to sequencing the drawing should represent the index of impairment rather than the end result. This largely represents the premise underlying the quantified process measures in the current study and the decision not to include measures of accuracy. Many of the existing qualitative scoring systems also give primacy to the sequence of drawing elements of the ROCF. In Bennett-Levy's (1984) system this is embedded in the concept of symmetry, whereby the order of executing ROCF elements is argued to reflect the individual's perceived structure and symmetry of the figure, that is, "successive construction of symmetrical units and their symmetrical components" (p. 113). Essentially what is highlighted is the individual's ability to extract the principle underlying how the figure is organised and sequence their construction accordingly.

The quantified process measures foreground temporal sequencing of visual details over spatial organisation as the primary measure of planning, specifically, planning strategy. In this context, fragmentation denotes poor planning strategy. This conceptual strategy imposes an appropriate framework for the copy of the ROCF. This is hypothesised to better identify strategies generating poor construction of the figure. In this context, the sequence of execution of component elements of the ROCF underlies the organisational approach to execution. Clinicians who routinely use the ROCF recognise that a variety of strategies can generate correct

and poor reproductions of the ROCF. This underscores that planning is a multidimensional construct, influenced by a range of cognitive mechanisms. The analysis of flowchart data in the study was undertaken well before reference was made to the final drawings and this revealed the divide between drawings that looked reasonably accurate but had been executed in a disjointed manner. This was more likely to be the case for normal subjects who might have deviated from an organised approach but were able to subsequently correct, even if incompletely, their drawings. In patients with acquired brain impairment, it is more likely the case that they are less able to compensate for poor initial approaches, which manifests in a poor copy of the ROCF. The weighted versions of the quantified process indices were ideally intended to reflect the importance of the initial stage of the temporal planning process, but it would appear that this wasn't reflected in superior psychometric outcomes for these measures. In other words, the initial stages in temporal planning were not associated with superior copy performances. This suggests that successful performance on the ROCF is contributed by many factors at different stages during copy of the figure.

There are of course disadvantages associated with choosing not to focus on what was drawn during copy trial of the ROCF but rather on how it was drawn. Drawing accuracy is considered important, driving the copy accuracy scoring system which has remained the hallmark of ROCF copy analysis. This is why many of the existing qualitative scoring systems include measures that reflect both how the drawing was executed and the accuracy of what was drawn, i.e. Hamby et al's (1993) scoring system and Stern et al's (1994) Boston Qualitative Scoring System. Bylsma's scoring system also required the investigator to visually inspect drawings additional to flowchart analysis since accuracy of some component elements of the ROCF is

included in the Q-Score. An interesting finding in the current study was that accuracy copy scores were significantly correlated with recall. A combination of sequencing and accuracy might therefore represent the appropriate middle ground for qualitative scoring systems for the ROCF. This direction will not be the focus of future research planned for quantified processing measures developed in the current study, where temporal sequencing of elements comprising the ROCF will continue to represent the primary area of focus.

Empirical, data-driven approaches have attracted criticism, largely because of the challenges in ascribing meaning to the results obtained. The guiding principle for data evaluation for novel process measures in the current study was that these characterised proximity of individual copy strategies to the normal reference group, with lower scores reflecting the degree of deviation from the reference group. Unlike the existing theory-driven scoring systems examined, there was no a priori hypothesis about an optimal copy strategy. While the caveats of data-driven approaches need to be acknowledged, one advantage of the current empirical approach is that unlike theory-driven scoring approaches, the chances of incorrectly classifying "normal" approaches as "abnormal" are minimised. This is of significant benefit in clinical settings where determinations around impaired ability guide diagnosis and clinical management.

One of the major difficulties with all the qualitative scoring approaches examined is that clinicians have been reluctant to use qualitative systems to score drawing approach for the ROCF. Additional time spent scoring drawing approach would appear to be the common complaint identified. This is unfortunate given the

important information provided by each of the qualitative scoring approaches examined. One of the advantages offered by the quantified process measures is that it is envisaged that at a future stage, information might be made available online such that clinicians can provide information regarding drawing approach for each client assessed, and output will be provided regarding the scores for nominated qualitative scoring systems, including the quantified process measures. Clinicians have also demonstrated poor uptake of existing qualitative scoring systems because it has been difficult to confidently appraise a multitude of approaches, each appearing to target different performance parameters.

While there is no question that examining qualitative aspects of performance on the ROCF can yield invaluable information about cognitive mechanisms underlying performance, what has been debated is what should be examined. Research on qualitative aspects of performance on the ROCF has been limited by the focus on diverse theoretical constructs underlying performance. While it is widely agreed that we should be looking more systematically at the how, there is less agreement about what it is we should be looking at. This in turn has created a vast number of research questions which have received little attention. Clinicians are therefore faced with the task of choosing scoring systems that have undergone limited validation. Clearly all of the qualitative scoring systems reviewed in the current study have merit. Findings from the current study add weight to a push for moving away from theoretical constructs defining the best approach to constructing the ROCF to empirical approaches which potentially better define expected performance in the context of relevant demographic variables. In the clinical context, concerns have been identified in regard to overstating cognitive impairment on the basis of limited test

data. It is imperative then to establish a reasonable level of confidence about the underlying mechanisms that are being assessed, and the nature of performance deviations that are assumed to denote impaired function. It is hoped that further research on qualitative assessment of performance on the ROCF gives clinicians and researchers a stronger understanding of the association between poor scores and cognitive impairment.

Strengths of the Current Study

Like many of the existing qualitative scoring systems for the ROCF, the current study endeavoured to draw on the tenets of process driven approaches to neuropsychological assessment. The point of departure from these scoring systems is represented by a move away from theoretically driven assumptions about optimal performance on the ROCF, leading instead to the establishment of empirically derived scoring measures which it is argued, better define optimal performance based on a normal reference group. Considerable attention was given as to how the construct of planning might be translated to copy approach for the ROCF, and how best to compare the derived quantified process measures with existing scoring approaches. There was extensive discussion between investigators regarding scoring parameters, contributing to the excellent inter-rater reliability across all study participants. The study investigator was blinded to the aetiology of acquired brain impairment for clinical participants at the time of flowchart analysis, reducing the potential for bias at this time. The absence of accuracy determinations in this novel scoring approach further reduced the scope for subjective interpretation of the appearance of individual elements, contributing to the reliability of scores obtained. One of the biggest advantages offered by the quantified process measures developed is that there was no

ambiguity as to how the index was derived given that once entered for analysis, flowchart data was programmed identically for conversion to the quantified process indices for all participants.

The importance of accurately representing each of the existing qualitative scoring systems was emphasised in the current study and translation of these scoring approaches was not programmed for analysis until there was 100% agreement between the study investigator and a co-investigator on all aspects of each of these scoring systems. Translating each of the qualitative scoring systems was a detailed, lengthy process. Some systems were easier to interpret than others. The most difficult system to translate was Bennett-Levy's (1984) Strategy Total score. As already noted, while the Good Continuation score was readily obtained, it was difficult to deduce the Symmetry Score. Despite this, every effort was made to ensure that the computerised scoring approaches for all of the existing qualitative scoring systems best represented the scoring methods intended by the original authors.

Limitations of the Current Study

An identified area of difficulty in the current study related to the use of the digital notepad (Digimemo system) for recording ROCF copy reproductions. Many drawings weren't captured due to patients not applying sufficient pressure whilst executing their copy of the ROCF. Therefore, in many cases, while the drawing was completely drawn on the paper attached to the digital notepad, this was not adequately captured digitally. All drawings where the digital copy was poorly captured or where there was considerable ambiguity regarding element execution,

were excluded from the study (12 cases from Full Normal group, 3 cases from Re-Test group, 21 cases from Clinical group). Significant efforts were made to thoroughly examine each of the digitally copied drawings in order to ensure that individual flowcharts across the study groups provided an exact representation of how subjects executed their copy. Future research might be undertaken using a more reliable digital system to capture the drawing process both on copy and on recall trials. While video recording did capture all aspects of the drawing in the Wilson & Batchelor (2015) study of undergraduate students, a large percentage of whom were included in the current study in the Wilson Normal group, clinically it is not always practical or possible to record subjects during assessment. The additional time required to undertake this task also makes it less attractive for busy clinicians. Ideally, the copy of the ROCF should be undertaken on a tablet where there is no need to transpose drawings and the data captured can be directly used for analysis. It is important to highlight, however, that the examiner will always be required to transcribe data captured digitally, particularly if sequencing measures represent the primary mode of analysis. In the current study, it was at times difficult to identify individual ROCF elements until the next element was drawn. This was particularly relevant for the left-hand vertical of the central rectangle, which is often continued to also form the left-hand vertical of the small rectangle below the central rectangle. Examiner input will always then be required to make judgments about which element is being scored.

The omission of a significant number of clinical cases, 21 cases in total, resulted in a heterogeneous clinical sample, without sufficient power to examine performance differences at a subgroup level. While the heterogeneity of the clinical

group was associated with increased variability in performance, the significant differences between clinical and normal participants across most scoring indices speaks to the strong effect of clinical status on copy approach to the ROCF overall. It might also be argued that this pattern of heterogeneity parallels the nature of referrals across a number of clinical settings. Future research should, however, target greater numbers of specific clinical groups to better ascertain differences in planning strategy on the ROCF copy across clinical subgroups.

Another limitation of the current study was the relatively small sample sizes used, especially in relation to the samples used to derive and validate the quantified process measures. Further studies are planned such that sufficient data can be collected across a wide range of age bands, enabling more accurate delineation of performance on the ROCF. It is of note that even in a relatively small sample size, widely reported age trends emerge, with performance on the ROCF declining as age increases. It is hoped that over time, enough data can be collected to enable a better understanding of performance parameters and their relationship to relevant demographic variables. Larger sample sizes will also facilitate an enhanced understanding of the stability of index scores over repeated administrations of the ROCF, which represents an integral part of neuropsychological assessment in the clinical context.

The current study did not investigate issues around convergent and divergent validity. Future research should ideally examine how derived quantified process measures relate to other executive measures, specifically measures of planning. This will be of integral importance in establishing that quantified process

measures do in fact represent a temporal measure of planning manifest by the sequence order in reproductions of the ROCF. This can only serve to enhance the potential appeal of these novel measures in offering clinicians information about planning approach to the ROCF, which can then be integrated with other measures of executive/adaptive function.

Conclusion

The novel approach adopted in the current study to develop quantified process measures for assessing planning on the ROCF establishes an important foundation for future research in this area. The quantified process measures developed compared favourably with existing qualitative scoring systems in producing scores which were normally distributed and further generated a wide range of scores across both normal and clinical participants. While reliability coefficients relating to consistency of scores on re-test did not perform favourably for these measures relative to existing qualitative scoring systems, quantified process measures were comparable to these systems in contributing significantly to the prediction of clinical status, and were related to recall of the ROCF. Collectively, these results are encouraging and support future research planned for further validation of these indices using larger sample sizes. Of all the quantified process indices developed, BAC would appear to be strongest measure in relation to the psychometric parameters investigated in the current study. It will be important to establish whether this remains the case in future studies, or whether in fact in a larger sample, the other quantified process indices might equally demonstrate strengths psychometrically in areas targeted for focus. Rather than simply adding another scoring system to the various qualitative scoring approaches currently available for the ROCF, it is hoped that the development and

validation of quantified process measures in the current study will encourage clinicians to increasingly consider available qualitative scoring systems to determine suboptimal performance on the ROCF. There is certainly much scope for improvement in the uptake of these scoring systems by clinicians.

References

Akshoomoff, N. A., & Stiles, J. (1995). Developmental trends in visuospatial analysis: Copying a complex figure. *Neuropsychology,* 9, 364-377.

Anderson, P., Anderson, V., & Garth, J. (2001). Assessment and development of organisational ability: The Rey Complex Figure Organisational Strategy Score (RCF-OSS). *Clinical Neuropsychology,* 15, 81-94.

Andres, P., & Van der Linden, M. (2001). Supervisory attentional system in patients with focal frontal lesions. *Journal of Clinical and Experimental Neuropsychology*, 23, 225-239.

Anzai, Y., & Simon, H. (1979). The theory of learning by doing. *Psychological Reviews*, 86, 124-140.

Ardila, A., & Rosselli, M. (1989). Neuropsychological characteristics of normal aging. *Developmental Neuropsychology,* 5, 307-320.

Ardila, A., Roselli, M., & Rosas, P. (1989). Neuropsychological assessment in illiterates: Visuospatial and memory abilities. *Brain and Cognition*, 11, 147-166.

Ashton, V. L., Donders, J., & Hoffman, N. M. (2005). Rey Complex Figure Test performance after traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology,* 27, 55-64.

Bennett-Levy, J. (1984). Determinants of performance on the Rey-Osterrieth Complex Figure Test: An analysis, and a new technique for single-case assessment. *British Journal of Clinical Psychology,* 23, 109-119.

Bernstein, J. H., & Waber, D. P. (1996). *Developmental Scoring System for the Rey-Osterrieth Complex Figure: Professional Manual.* Lutz FL: Psychological Assessment Resources.

Berry, D. T. R., Allen, R. S., & Schmitt, F. A. (1991). Rey-Osterrieth Complex Figure: Psychometric characteristics in a geriatric sample. The *Clinical Neuropsychologist,* 5, 143-153.

Berry, D. T. R., & Carpenter, G. S. (1992). Effect of four different delay periods on recall of the Rey-Osterrieth Complex Figure by older persons. *The Clinical Neuropsychologist,* 6, 80-84.

Binder, L. (1982). Constructional strategies on complex figure drawings after unilateral brain damage. *Journal of Clinical Neuropsychology,* 4, 51-58.

Boone, K. B., Lesser, I. M., Hill-Gutierrez, E., Berman, N., & D'Elia, L. F. (1993). Rey-Osterrieth Complex Figure performance in healthy, older adults: Relationship to age, education, sex and IQ. *The Clinical Neuropsychologist,* 7, 22-28.

Breier, J. I., Plenger, P. M., Castillo, R., Fuchs, K., Wheless, J. W., Thomas, A. B., Brookshire, B. L., Willmore, L. J., & Papanicolaou, A. (1996). Effects of temporal lobe epilepsy on spatial and figural aspects of memory for a complex geometric figure. *Journal of the International Neuropsychological Society,* 2, 535-540.

Buhlmann, U., Deckersbach, T., Engelhard, I., Cook, L. M., Rauch, S. L., Kathmann, N., Wilhelm, S., & Savage, C. (2006). Cognitive retraining for organisational impairment in obsessive-compulsive disorder. *Psychiatry Research,* 144, 109-116.

Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton Tests.* Thurston, Suffolk: Thames Valley Test Company.

Bylsma, F. W., Bobholz, J. H., Schretlen, D., & Correa, D. D. (1995). A brief, reliable approach to coding how subjects copy the Rey-Osterrieth Complex Figure (Abstract). *Journal of the International Neuropsychological Society,* 1, 125.

Bylsma, F. W., Carlson, M. C., Schretlen, D., Zonderman, A., & Resnick, S. (1997). Rey-Osterrieth Complex Figure Test (CFT) Q-Score performance in 328 healthy adults ages 20 to 94 (Abstract). *Journal of the International Neuropsychological Society,* 3, 70.

Bylsma, F. W. (2008). *The Q-Score: A brief reliable method for coding how subjects copy the Rey-Osterrieth Complex Figure.* Chicago, IL: Neuropsychological Services.

Carey, C. L., Woods, S. P., Damon, J., Halabi, C., Dean, D., Delis, D., Miller, B. L., & Kramer, J. H. (2008). Discriminant validity and neuroanatomical correlates of rule monitoring in frontotemporal dementia and Alzheimer's disease. *Neuropsychologia,* 46, 1081-1087.

Carlin, D., Bonerba, J., Phipps, M., Alexander, G., Shapiro, M., & Grafman, J. (2000). Planning impairments in frontal lobe dementia and frontal lobe lesion patients. *Neuropsychologia,* 38, 655-665.

Carr, E. K., & Lincoln, N. B. (1988). Interrater reliability of the Rey figure copying test. *British Journal of Clinical Psychology,* 27, 267-268.

Casey, M. B., Winner, E., Hurwitz, I., & DaSilva, D. (1991). Does processing style affect recall of the Rey-Osterrieth or Taylor Complex Figures. *Journal of Clinical and Experimental Neuropsychology,* 13, 600-606.

Chiulli, S. J., Haaland, K. Y., LaRue, A., & Garry, P. J. (1995). Impact of age on drawing the Rey-Osterrieth figure. *The Clinical Neuropsychologist,* 9, 219-224.

Cockburn, J. (1995). Performance on the Tower of London test after severe head injury. *Journal of the International Neuropsychological Society,* 1, 537-544.

Corwin, J., & Bylsma, F. W., (1993). Translations of excerpts from Andre Rey's psychological examination of traumatic encephalopathy and P. A. Osterrieth's The Complex Figure Test. *The Clinical Neuropsychologist,* 7, 3-15.

Dawson, L. K., & Grant, I. (2000). Alcoholics' initial organisational and problem-solving skills predict learning and memory performance on the Rey-Osterrieth Complex Figure. *Journal of the International Neuropsychological Society,* 6, 12-19.

Deckersbach, T., Savage, C. R., Phillips, C. A., Wilhelm, S., Bulman, U., Rauch, S. L., Baer, L., & Jenike, M. A. (2000). Reliability and validity of a scoring system for measuring organisational approach in the Complex Figure Test. *Journal of Clinical and Experimental Neuropsychology,* 22, 640-648.

Delaney, R. C., Prevey, M. L., Cramer, J., Mattson, R. H., & VA Epilepsy Co-operative Study # 264 Research Group. (1992). Test-retest comparability and control subject data for the Rey Auditory Verbal Learning Test and the Rey-Osterrieth/Taylor Complex Figures. *Archives of Clinical Neuropsychology*, 7, 523-528.

Delis, D. C. (1989). Neuropsychological assessment of learning and memory. In F. F. Boller & J. Grafman (Eds.), *Handbook of Neuropsychology* (Vol 3, 3-33). Amsterdam: Elsevier.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *The Delis-Kaplan Executive Function System: Technical Manual.* San Antonio, TX: The Psychological Corporation.

Denman, S. B. (1984). *Denman Neuropsychology Memory Scale.* Charleston, S. C: Sidney B. Denman.

Diamond. B. J., DeLuca, J., & Kelley, S. M. (1997). Memory and executive functions in amnesic and non-amnesic patients with aneurysms of the anterior communicating artery. *Brain,* 12, 1015-1025.

Duley, J. F., Wilkins, J. W., Hamby, S. L., Hopkins, D. G., Burwell, R. D., & Barry, N. S. (1993). Explicit scoring criteria for the Rey-Osterrieth and Taylor Complex Figures. The *Clinical Neuropsychologist,* 7, 29-38.

Erickson, R. C. (1995). A review and critique of the process approach in neuropsychological assessment. *Neuropsychology Review,* 5, 223-243.

Fastenau, P. S., Denburg, N. L., & Hufford, B. J. (1999). Adult norms for the Rey-Osterrieth Complex Figure Test and for supplemental recognition and matching trials from the Extended Complex Figure Test. *The Clinical Neuropsychologist,*13, 30-47.

Ferraro, F. R., Grossman, J., Bren, A., & Hoverson, A. (2002). Effects of orientation on Rey Complex Figure performance. *Brain and Cognition,* 50, 139-144.

Gallagher, C., & Burke, T. (2007). Age, gender and IQ effects on the Rey-Osterrieth Complex Figure Test. *British Journal of Clinical Psychology,* 46, 35-45.

Grossman, M., Carvell, S., Peltzer, L., Stern, M., Gollomp, S., & Hurtig, H. (1993) Visual construction impairments in Parkinson's Disease. *Neuropsychology*, 7, 536-547.

Hamby, S. L., Wilkins, J. W., & Barry, N. S. (1993). Organisational quality on the Rey-Osterrieth and Taylor Complex Figure Tests: A new scoring system. *Psychological Assessment,* 5, 27-33.

Hartman, M., & Potter, G. (1998). Sources of age differences on the Rey-Osterrieth Complex Figure Test. *The Clinical Neuropsychologist,* 12, 513-524.

Humes, G. E., Welsh, M. C., Retzlaff, P., & Cookson, N. (1997). Towers of Hanoi and London: Reliability of two executive function tasks. *Assessment,* 4, 249-257.

Janowsky, J. S., & Thomas-Thrapp, L. J. (1993). Complex Figure recall in the elderly: A deficit in memory or constructional strategy. *Journal of Clinical and Experimental Neuropsychology*, 15, 159-169.

Kafer, K. L., & Hunter, M. (1997). On testing the face validity of planning/problem solving tasks in a normal population. *Journal of the International Neuropsychological Society*, 3, 108-119.

Kaller, C. P., Unterrainer, J. M., Rahm, B., & Halsband, U. (2004). The impact of problem structure on planning: Insights from the Tower of London task. *Cognitive Brain Research,* 20, 462-472.

Kaplan, E. (1983). Process and achievement revisited. In S. Waper & B. Kaplan (Eds.), *Toward a Holistic Development of Psychology.* Hillsdale, N. J.

Erlbaum.

Kaplan, E. (1988). A process approach to neuropsychological assessment. In T. Boll & B. K. Bryant (Eds.), *Clinical Neuropsychology and Brain Function: Research, Measurement, and Practice.* Washington, DC: American Psychological Association.

Kaplan, E., Fein, D., Morris, R., & Delis, D. (1991). *WAIS-R NI Manual.* San Antonio, TX: The Psychological Corporation.

King, M. C. (1981). Effects of non-focal brain dysfunction on visual memory. *Journal of Clinical Psychology*, 37, 638-643.

Kixmiller, J. S., Verfaellie, M., Mather, M. M., & Cermak, L. S. (2000). Role of perceptual and organisational factors in amnesics' recall of the Rey Osterrieth Complex Figure: A comparison of three amnesic groups. *Journal of Clinical and Experimental Neuropsychology*, 22, 198-207.

Knight, J. A. (2003a). The Rey-Osterrieth Complex Figure: Overview of the handbook, current uses and future directions. In *The Handbook of Rey-Osterrieth Complex Figure Usage: Clinical and Research Applications*. Lutz, Fl: Psychological Assessment Resources.

Knight, J. A. (2003b). ROCF administration procedures and scoring systems. In *The Handbook of Rey-Osterrieth Complex Figure Usage: Clinical and Research Applications.* Knight, J. A. (Ed.). Lutz, Fl: Psychological Assessment Resources.

Krikorian, R., Bartok, J., & Gay, N. (1994). Tower of London procedure: A standard method and developmental data. *Journal of Clinical and Experimental Neuropsychology*, 16, 840-850.

Lange, G., Waked, W., Kirshblum, S., & DeLuca, J. (2000). Organisational strategy influence on visual memory performance after stroke, cortical/ subcortical and left/right hemisphere contrasts. *Archives of Physical Medicine and Rehabilitation,* 1, 89-94.

Levine, A. J., Miller, E. N., Becker, J. T, Selnes, O. A., & Cohen, B. A. (2004). Normative data for determining significance of test-retest differences on eight common neuropsychological instruments. *The Clinical Neuropsychologist,* 18, 373-384.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological Assessment.* New York: Oxford University Press.

Liberman, J., Stewart, W., Seines, O., & Godon, B. (1994). Rater agreement for the Rey-Osterrieth Complex Figure Test. *Journal of Clinical Psychology,* 50, 615-624.

Loring, D. W., Martin, R. C., Meador, K. J., & Lee, G. P. (1990). Psychometric construction of the Rey-Osterrieth Complex Figure: Methodological considerations and interrater reliability. *Archives of Clinical Neuropsychology,* 5, 1-14.

Maillet, D. J. (1992). Copying strategy on the Rey-Osterrieth Complex Figure Test: Performance in normal adults. Unpublished Masters Thesis. Saint Mary's University: Halifax.

Meyers, L., & Meyers, K. R. (1995). *Rey Complex Figure Test and Recognition Trial.* Odessa, FL: Psychological Assessment Resources.

Milberg, W. P. & Hebben, N. (2006). The historical antecedents of the Boston Process Approach. In A. M. Poreh (Ed.), *The Quantified Process Approach to Neuropsychological Assessment.* New York: Taylor &

Francis.

Mitrushina, M., & Satz, P. (1991). Effect of repeated administration of a neuropsychological battery in the elderly. *Journal of Clinical Psychology*, 47, 790-801.

Mitrushina, M., Satz, P., & Chervinsky (1990). Efficiency of recall on the Rey-Osterrieth Complex Figure in normal aging. *Brain Dysfunction,* 3, 148-150.

Morris, R. G., Downes, J. J., Sahakian, B. J., Evenden, J. L., Heald, A. & Robbins, T. W. (1988). Planning and spatial working memory in Parkinson's Disease. *Journal of Neurology, Neurosurgery and Psychiatry,* 51, 757-766.

Morris, R. G., Ahmed, S., Syed, G. M., & Toone, B. K. (1994). Neural correlates of Planning ability: Frontal lobe activation during the Tower of London Test. *Neuropsychologia,* 31, 1367-1378.

Norman, D. A., & Shallice, T. (1986). Attention to action. Willed and automatic control of behaviour. In *Consciousness and Self-Regulation*. Vol. 4. Davidson, R. J., Schwartz, G. & Shapiro, D. (Eds.). New York: Plenum Press.

Osterrieth, P. A. (1944). Le test de copie d'une figure complex: Contribution a l'etude de la perception et de la memoire [The test of copying a complex figure: Acontribution to the study of perception and memory] *Archives de Psychologie,* 30, 206-356.

Ostrosky-Solis, F., Jaime, R. M., & Ardila, A. (1998). Memory abilities during normal aging. *International Journal of Neuroscience,* 93, 151-162.

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia,* 28, 1021-1034.

Penades, R., Catalan, R., Andres, S., Salamero, M., & Gasto, C. (2005). Executive function and nonverbal memory in obsessive-compulsive disorder. *Psychiatry Research,* 133, 81-90.

Phillips, L. H., Wynn, V., Gilhooly, K. J., Della Salla, S., & Logie, R. H. (1999). The role of memory in the Tower of London task. *Memory,* 7, 209-231.

Ponton, M. O., Satz, P., Herrera, L., Oritz, F., Urritia, C. P., Young, r., et al. (1996). Normative data stratified by age and education for the Neuropsychological Screening Battery for Hispanics (NeSBHIS): Initial report. *Journal of the International Neuropsychological Society,* 2, 96-104.

Poreh, A. M. (2000). The quantified process approach: An emerging methodology in neuropsychological assessment. *The Clinical Neuropsychologist,* 14, 212-222.

Poreh, A. M. (2006). A brief introduction to the Quantified Process Approach. In A. M. Poreh (Ed.), *The Quantified Process Approach to Neuropsychological Assessment.* New York: Taylor & Francis

Poreh, A. M., & Shye, S. (1998). Examination of the global and local features of the Rey Osterrieth Complex Figure using faceted smallest space analysis. *The Clinical Neuropsychologist, 12,* 453-467.

Powell, G. E. (1979). The relationship between intelligence and verbal and spatial memory. *Journal of Clinical Psychology,* 35, 335-340.

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical Neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, 20, 33-65.

Rainville, C., Amieva, H., LaFont, S., Dartigues, J. F., Orgogozo, J. M., & Fabrigoule, C. (2002). Executive function deficits in patients with dementia of the Alzheimer's type. A study with a Tower of London task. *Archives of Clinical Neuropsychology*, 17, 513-530.

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie tramatique (Les problemes) [Psychological examination in cases of traumatic encephalopathy]. *Archives de Psychologie*, 28, 286-340.

Rosselli, M., & Ardila, A. (1991). Effects of age, education, and gender on the Rey-Osterrrieth Complex Figure. *The Clinical Neuropsychologist*, 5, 370-376.

Savage, C. R., Baer, L., Keuthen, N. J., Brown, H. D., Rauch, S. L., & Jenike, M. A. (1999). Organisational strategies mediate nonverbal memory impairment in obsessive-compulsive disorder. *Biological Psychiatry*, 45, 905-916.

Savage, C. R., Deckersbach, T., Wilhelm, S., Rauch, S., Scott, L., Baer, L., Reid, T., & Jenike, M. A. (2000). Strategic processing and episodic memory impairment in obsessive compulsive disorder. *Neuropsychology*, 14, 141-151.

Schall, U., Johnston, P., Lagopoulos, J., Juptner, M., Jentzen, W., Thienel, R., Dittman-Balcar, A., Bender, S., & Ward, P. B. (2003). Functional Brain maps of Tower of London performance: A positron emission tomography and functional magnetic resonance imaging study. *Neuroimage,* 20, 1154-1161.

Seidman, L. J., Benedict, K. B., Biederman, J., Bernstein, J. H., Seivard, K., Milberger, S., Norman, D., Mick, F. & Faraone, V. (1995). Performance of children with ADHD on the Rey-Osterrieth Complex Figure: A Pilot neuropsychological study. *Journal of Child Psychology and Psychiatry,* 36, 1459-1473.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of The Royal Society of London,* B298, 199-209.

Shin, M. S., Park, S. J., Kim, M. S., Lee, Y. H., Ha, T. H., & Kwon, J. S. (2004). Deficits of organisational strategy and visual memory in obsessive-compulsive disorder. *Neuropsychology,* 18, 665-672.

Shin, N. Y., Kang, D. H., Choi, J. S, Jung, M. H, Jang, J. H, & Kwon, J. S. (2010). Do organisational strategies mediate nonverbal memory impairment in drug naïve patients with obsessive-compulsive disorder? *Neuropsychology*, 24, 527-533.

Shorr, J. S., Delis, D. C., & Massman, P. J. (1992). Memory for the Rey-Osterrieth Figure: Perceptual clustering, encoding, and storage. *Neuropsychology,* 6, 43-50.

Ska, B., & Nespoulous, J. (1988). Encoding strategies and recall performance of a complex figure by normal elderly subjects. *Canadian Journal on Aging,* 7, 408-416.

Smith, S. R., Chang, J., Shnoebelen, K. J., Edwards, J. W., Servesko, A. M., & Walker, S. J. (2007). Psychometrics of a simple method for scoring organisational approach on the Rey-Osterrieth Complex Figure. *Journal of Neuropsychology*, 1, 39-51.

Stern, R. A., Javorsky, D. J., Singer, E. A., Singer Harris, N. G., Somerville, J. A., Duke, L., Thompson, J. A., & Kaplan, E. (1999). *The Boston Qualitative Scoring System for the Rey-Osterrieth Figure: Professional Manual.* Lutz, FL: Psychological Assessment Resources.

Strauss, E., & Spreen, O. (1990). A comparison of the Rey and Taylor figures. *Archives of Clinical Neuropsychology*, 5, 417-420.

Sullivan, E. V., Mathalon, D. H., Ha, C. N., Zipursky, R. B., & Pfefferbaum, A. (1992). The contribution of constructional accuracy and organisational strategy to nonverbal recall in schizophrenia and chronic alcoholism. *Biological Psychiatry*, 32, 312-333.

Taylor, E. M. (1959). *Psychological Appraisal of Children with Cerebral Deficits.* Cambridge, M. A: Harvard University Press.

Taylor, L. B. (1969). Localisation of cerebral lesions by psychological testing. *Clinical Neurosurgery,* 16, 269-287.

Temple, R. O., Duncan Davis, J., Silverman, I., & Tremont, G. (2006). Differential impact of executive function on visual memory tasks. *The Clinical Neuropsychologist,* 20, 480-490.

Tombaugh, T. N. & Hubley, A. M. (1991) Four studies comparing the Rey-Osterrieth and Taylor Complex Figures. *Journal of Clinical and Experimental Neuropsychology*, 13, 587-599.

Tombaugh, T. N., Faulkner, P., & Hubley, A. M. (1992). Effects of age on the Rey-Osterrieth and Taylor Complex Figures: Test-retest data using an intentional learning paradigm. *Journal of Clinical and Experimental Neuropsychology*, 14, 647-661.

Troyer, A. K., & Wishart, H. A. (1997). A comparison of qualitative scoring systems for the Rey-Osterrieth Complex Figure Test. *The Clinical Neuropsychologist*, 11, 381-390.

Tupler, L. A., Welsh, K. A., Asare-Aboagye, Y., & Dawson, D. V. (1995). Reliability of the Rey-Osterrieth Complex Figure in use with memory impaired patients. *Journal of Clinical and Experimental Neuropsychology*, 17, 566-579.

Visser, R. S. H. (1973). *Manual of the Complex Figure Test*. Amsterdam: Swets & Zeitlinger.

Waber, D. P., & Holmes, J. M. (1985). Assessing children's copy productions of the Rey-Osterrieth Complex Figure. *Journal of Clinical and Experimental Neuropsychology*, 7, 264-280.

Waber, D. P., & Holmes, J. M. (1986). Assessing children's memory productions of the Rey-Osterrieth Complex Figure. *Journal of Clinical and Experimental Neuropsychology*, 8, 563-580.

Waber, D. P., Isquith, P. K., Kahn, C. M., Romero, I., Sallan, S. E., & Tarbell, N. J. (1994). Metacognitive factors in the visuospatial skills of long-term survivors of acute lymphoblastic leukaemia (ALL): An experimental approach to the Rey-Osterrieth Complex Figure Test. *Developmental Neuropsychology*, 10, 349-367.

Weider, S., Indredavik, M. S., Lydersen, S., & Hestad, K. (2016). Central coherence, visuoconstruction and visual memory in patients with eating disorders as measured by different scoring methods of the Rey Complex Figure Test. *European Eating Disorders Review,* 24, 106-113.

Wertheimer, M. (1958). Principles of perceptual organisation. In D. C. Beardslee & M. Wertheimer (Eds), *Readings in Perception.* Princeton, NJ: Van Nostrand.

Westervelt, H. J., Somerville, J., Tremont, G., & Stern, R. A. (2000). The impact of organisational strategy on recall of the Rey-Osterrieth Complex Figure. (Abstract). *Archives of Clinical Neuropsychology,* 15, 684.

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). Behavioural Assessment of the Dysexecutive Syndrome

Wilson, N., & Batchelor, J. (2015). Examining Rey Complex Figure Test organisation in healthy adults. *Journal of Clinical and Experimental Neuropsychology*, 37, 1052-1061.

Yeates, K. O., Patterson, C. M., Waber, D. P., & Holmes Bernstein, J. (2003). Constructional and figural memory skills following paediatric closed head injury: Evaluation using the ROCF. In *The Handbook of Rey-Osterrieth Complex Figure Usage: Clinical and Research Applications.* Knight, J. A. (Ed.). Lutz, FL: Psychological Assessment Resources.

APPENDICES

Appendix 1 of this thesis has been removed as it may contain sensitive/confidential content

Appendix 2: Proportion Tables for Each of the Novel Process Measures

Table 10. *Proportion of the Reference Normal Sample Completing Each of the 18 ROCF Elements (Vertical Axis) by Order (Horizontal Axis) - BAC*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0.3	0.1	0	0	0	0	0	0	0	0	0.04	0	0	0	0.04	0.13	0.08	0.21
2	0.5	0	0	0	0	0.1	0	0.1	0	0.08	0.04	0	0.04	0.04	0.04	0	0	0
3	0	0	0	0	0.1	0	0.1	0	0.1	0.13	0.08	0.08	0.08	0.04	0.08	0.08	0.04	0
4	0	0.1	0.1	0.1	0	0	0	0.2	0.1	0.08	0.08	0	0.08	0	0	0	0	0
5	0	0.2	0.1	0.2	0.2	0.1	0	0	0	0.08	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0.1	0	0	0	0	0	0.17	0.04	0.17	0.13	0.21	0.04
7	0	0	0	0	0	0.1	0	0	0	0	0	0.08	0	0.04	0.08	0.21	0.04	0.08
8	0	0	0	0	0.1	0	0	0.1	0	0.04	0.08	0.13	0.04	0.13	0.13	0	0.13	0.04
9	0	0.2	0.1	0.1	0.2	0.1	0	0	0.1	0.04	0	0.04	0	0.04	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0.17	0.13	0.17	0.17	0.08	0.04	0.04	0.04	0
11	0	0	0	0	0	0	0	0	0	0	0.08	0.17	0.25	0.13	0.04	0.13	0.04	0.04
12	0	0	0	0	0	0	0	0	0.1	0.04	0.13	0.21	0.04	0.25	0.08	0.04	0.08	0
13	0.1	0.2	0	0.1	0	0.3	0	0	0	0.04	0	0	0	0	0.04	0.04	0	0
14	0	0	0.2	0	0.1	0	0.2	0.1	0	0.08	0	0	0.04	0	0	0	0.08	0.13
15	0	0	0	0	0	0.2	0.2	0	0	0.08	0.04	0	0.04	0.04	0	0.04	0.13	0.04
16	0	0	0.1	0	0.2	0	0.1	0.1	0.2	0.04	0.08	0.04	0	0	0.04	0.04	0	0
17	0	0	0.1	0	0	0	0.1	0.1	0	0.04	0.21	0	0.04	0	0.08	0.08	0.04	0.04
18	0	0.2	0.1	0.1	0	0	0	0	0.1	0.04	0	0.08	0	0.17	0.08	0	0.04	0

Table 11. *Weighted Proportion of the Reference Normal Sample Completing Each of the 18 ROCF Elements by Order – BAC-W*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**4.5**	**1.4**	**0.7**	**0**	**0.6**	**0**	**0**	**0.5**	**0.4**	**0**	**0.33**	**0**	**0**	**0**	**0.17**	**0.38**	**0.17**	**0.21**
2	**9.75**	**0**	**0**	**0**	**0**	**1.1**	**0**	**0.9**	**0.4**	**0.75**	**0.33**	**0**	**0.25**	**0.21**	**0.17**	**0**	**0**	**0**
3	**0**	**0**	**0.7**	**0.6**	**1.2**	**0**	**1.0**	**0.5**	**0.8**	**1.13**	**0.67**	**0.58**	**0.5**	**0.21**	**0.33**	**0.25**	**0.08**	**0**
4	**0.75**	**1.4**	**1.3**	**1.9**	**0**	**0.5**	**0.5**	**1.8**	**1.3**	**0.75**	**0.67**	**0**	**0.5**	**0**	**0**	**0**	**0**	**0**
5	**0.75**	**2.8**	**2.0**	**3.1**	**2.3**	**1.1**	**0.5**	**0.5**	**0.4**	**0.75**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0.7**	**0**	**0**	**0**	**1.0**	**0**	**0**	**0**	**0**	**0**	**1.0**	**0.21**	**0.67**	**0.38**	**0.42**	**0.04**
7	**0**	**0**	**0**	**0.6**	**0.6**	**1.1**	**0**	**0**	**0.4**	**0**	**0**	**0.58**	**0**	**0.21**	**0.33**	**0.63**	**0.08**	**0.08**
8	**0**	**0**	**0**	**0.6**	**1.2**	**0.5**	**0**	**0.9**	**0.4**	**0.38**	**0.67**	**0.88**	**0.25**	**0.63**	**0.5**	**0**	**0.25**	**0.04**
9	**0.75**	**2.8**	**2.0**	**1.9**	**2.3**	**1.1**	**0.5**	**0**	**1.3**	**0.38**	**0**	**0.29**	**0**	**0.21**	**0**	**0**	**0**	**0**
10	**0**	**0**	**0**	**0.6**	**0**	**0.5**	**0**	**0.5**	**0.4**	**1.5**	**1.0**	**1.17**	**1.0**	**0.42**	**0.17**	**0.13**	**0.08**	**0**
11	**0**	**0**	**0**	**0**	**0.6**	**0**	**0.5**	**0.5**	**0**	**0**	**0.67**	**1.17**	**1.5**	**0.63**	**0.17**	**0.38**	**0.08**	**0.04**
12	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0.5**	**0.8**	**0.38**	**1.0**	**1.46**	**0.25**	**1.25**	**0.33**	**0.13**	**0.17**	**0**
13	**1.5**	**3.5**	**0.7**	**1.9**	**0.6**	**3.8**	**0.5**	**0.5**	**0**	**0.38**	**0**	**0**	**0**	**0**	**0.17**	**0.13**	**0**	**0**
14	**0**	**0**	**2.7**	**0.6**	**1.8**	**0**	**2.0**	**1.4**	**0.4**	**0.75**	**0**	**0**	**0.25**	**0**	**0**	**0**	**0.17**	**0.13**
15	**0**	**0**	**0.7**	**0.6**	**0.6**	**2.2**	**2.0**	**0.5**	**0.4**	**0.75**	**0.33**	**0**	**0.25**	**0.21**	**0**	**0.13**	**0.25**	**0.04**
16	**0**	**0.7**	**1.3**	**0.6**	**2.3**	**0**	**1.5**	**0.9**	**1.7**	**0.38**	**0.67**	**0.29**	**0**	**0**	**0.17**	**0.13**	**0**	**0**
17	**0**	**0.7**	**2.0**	**0.6**	**0**	**0.5**	**1.5**	**0.9**	**0**	**0.38**	**1.67**	**0**	**0.25**	**0**	**0.33**	**0.25**	**0.08**	**0.04**
18	**0**	**3.5**	**1.3**	**1.3**	**0**	**0.5**	**0.5**	**0.5**	**0.8**	**0.38**	**0**	**0.58**	**0**	**0.83**	**0.33**	**0**	**0.08**	**0**

Table 12. *Proportion of the Reference Normal Sample Attempting Each of the 18 ROCF Elements by Order - BAA*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**0.3**	**0.1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0.04**	**0**	**0**	**0**	**0**	**0.13**	**0.13**	**0.17**
2	**0.6**	**0.3**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
3	**0**	**0**	**0**	**0.1**	**0**	**0.1**	**0.1**	**0.1**	**0.2**	**0.17**	**0.13**	**0.08**	**0**	**0**	**0**	**0**	**0**	**0**
4	**0**	**0.1**	**0**	**0.1**	**0**	**0.1**	**0.1**	**0.2**	**0.1**	**0.08**	**0.04**	**0.04**	**0.04**	**0**	**0**	**0**	**0**	**0**
5	**0**	**0**	**0.1**	**0.1**	**0.3**	**0.1**	**0.1**	**0**	**0**	**0.08**	**0.04**	**0**	**0.04**	**0**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0.08**	**0.08**	**0.04**	**0**	**0.13**	**0.21**	**0.13**	**0.21**	**0.04**
7	**0**	**0**	**0**	**0**	**0**	**0**	**0.1**	**0**	**0**	**0**	**0.04**	**0**	**0.08**	**0**	**0.04**	**0.21**	**0.13**	**0.08**
8	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0.1**	**0**	**0.13**	**0.04**	**0.08**	**0.04**	**0.17**	**0.08**	**0.08**	**0.08**	**0.08**
9	**0**	**0.1**	**0**	**0.2**	**0.2**	**0.2**	**0**	**0**	**0**	**0.04**	**0.08**	**0.04**	**0**	**0.04**	**0**	**0**	**0**	**0**
10	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0.08**	**0.17**	**0.21**	**0.13**	**0.08**	**0.13**	**0.08**	**0**	**0.04**
11	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0.04**	**0.04**	**0.13**	**0.25**	**0.13**	**0.13**	**0.13**	**0.04**	**0.08**
12	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0.08**	**0.17**	**0.21**	**0.25**	**0.08**	**0.04**	**0.08**	**0.04**
13	**0**	**0.1**	**0.3**	**0**	**0.1**	**0.1**	**0.2**	**0**	**0**	**0**	**0**	**0.08**	**0**	**0**	**0.04**	**0.04**	**0**	**0**
14	**0**	**0**	**0.1**	**0.1**	**0**	**0.1**	**0.1**	**0**	**0.1**	**0.04**	**0.04**	**0.04**	**0.04**	**0**	**0.04**	**0**	**0.08**	**0.13**
15	**0**	**0**	**0**	**0**	**0.1**	**0.1**	**0**	**0.2**	**0**	**0.17**	**0.04**	**0**	**0.04**	**0**	**0**	**0**	**0.21**	**0.04**
16	**0**	**0**	**0.1**	**0**	**0**	**0.1**	**0**	**0.1**	**0.2**	**0**	**0.04**	**0.04**	**0.04**	**0.08**	**0.04**	**0.04**	**0**	**0**
17	**0**	**0**	**0**	**0.3**	**0**	**0**	**0.1**	**0**	**0**	**0.08**	**0.08**	**0**	**0.08**	**0**	**0.08**	**0.04**	**0**	**0**
18	**0**	**0.2**	**0.3**	**0**	**0**	**0**	**0**	**0**	**0.1**	**0**	**0**	**0.04**	**0**	**0.13**	**0.13**	**0.04**	**0**	**0.04**

Table 13. *Weighted Proportion of the Reference Normal Sample Attempting Each of the 18 ROCF Elements by Order – BAA-W*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**5.25**	**1.4**	**0.7**	**0**	**0**	**0**	**0.5**	**0.5**	**0.4**	**0**	**0.33**	**0**	**0**	**0**	**0**	**0.38**	**0.25**	**0.17**
2	**11.3**	**5.0**	**0**	**0**	**0**	**0.5**	**0.5**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
3	**0**	**0**	**0.7**	**1.3**	**0.6**	**1.1**	**1.5**	**0.9**	**1.7**	**1.5**	**1.0**	**0.58**	**0**	**0**	**0**	**0**	**0**	**0**
4	**0.75**	**1.4**	**0.7**	**1.3**	**0.6**	**1.6**	**1.0**	**1.8**	**1.3**	**0.75**	**0.33**	**0.29**	**0.25**	**0**	**0**	**0**	**0**	**0**
5	**0**	**0.7**	**2.0**	**1.3**	**4.1**	**1.6**	**1.5**	**0.5**	**0**	**0.75**	**0.33**	**0**	**0.25**	**0**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0**	**0**	**0.6**	**0.5**	**0**	**0**	**0**	**0.75**	**0.67**	**0.29**	**0**	**0.63**	**0.83**	**0.38**	**0.42**	**0.04**
7	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0.5**	**0.4**	**0**	**0.33**	**0**	**0.5**	**0**	**0.17**	**0.63**	**0.25**	**0.08**
8	**0**	**0**	**0**	**0**	**0.6**	**0**	**0.5**	**0.9**	**0.4**	**1.13**	**0.33**	**0.58**	**0.25**	**0.83**	**0.33**	**0.25**	**0.17**	**0.08**
9	**0.75**	**2.1**	**0.7**	**3.1**	**2.3**	**2.2**	**0**	**0**	**0.4**	**0.38**	**0.67**	**0.29**	**0**	**0.21**	**0**	**0**	**0**	**0**
10	**0**	**0**	**0**	**0**	**0.6**	**0**	**0**	**0**	**0.4**	**0.75**	**1.33**	**1.46**	**0.75**	**0.42**	**0.5**	**0.25**	**0**	**0.04**
11	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0.5**	**0**	**0.38**	**0.33**	**0.88**	**1.5**	**0.63**	**0.5**	**0.38**	**0.08**	**0.08**
12	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0.5**	**0**	**0**	**0.67**	**1.17**	**1.25**	**1.25**	**0.33**	**0.13**	**0.17**	**0.04**
13	**0**	**2.1**	**4.0**	**0.6**	**1.8**	**1.1**	**2.0**	**0**	**0.4**	**0**	**0**	**0.58**	**0**	**0**	**0.17**	**0.13**	**0**	**0**
14	**0**	**0**	**1.3**	**1.3**	**0.6**	**1.1**	**1.5**	**0.5**	**1.3**	**0.38**	**0.33**	**0.29**	**0.25**	**0**	**0.17**	**0**	**0.17**	**0.13**
15	**0**	**0**	**0**	**0.6**	**1.2**	**1.1**	**0.5**	**2.3**	**0**	**1.5**	**0.33**	**0**	**0.25**	**0**	**0**	**0**	**0.42**	**0.04**
16	**0**	**0.7**	**1.3**	**0**	**0.6**	**1.6**	**0.5**	**1.4**	**2.1**	**0**	**0.33**	**0.29**	**0.25**	**0.42**	**0.17**	**0.13**	**0**	**0**
17	**0**	**0**	**0.7**	**5.0**	**0.6**	**0.5**	**1.0**	**0.5**	**0.4**	**0.75**	**0.67**	**0**	**0.5**	**0**	**0.33**	**0.13**	**0**	**0**
18	**0**	**3.5**	**4.0**	**0.6**	**0**	**0**	**0**	**0.5**	**0.8**	**0**	**0**	**0.29**	**0**	**0.63**	**0.5**	**0.13**	**0**	**0.04**

Appendix 3: Frequency Histograms for Wilson Normal Group Across Existing Qualitative Scoring Systems and Novel Process Measures

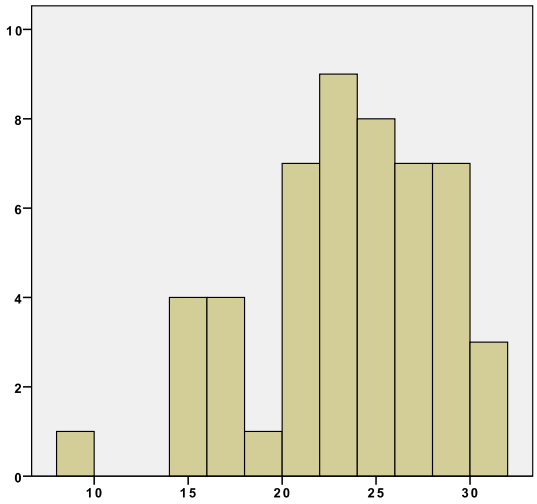

Figure 38. Frequency histogram of Bennett-Levy scores for the Wilson Normal group (N=51). Vertical axis shows frequencies out of 10 and horizontal axis shows scores.

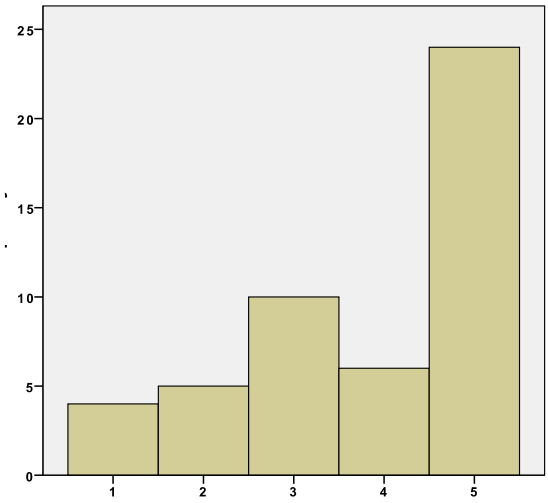

Figure 39. Frequency histogram of Binder scores for the Wilson Normal group (N=51). Vertical axis shows frequencies out of 25 and horizontal axis shows scores.

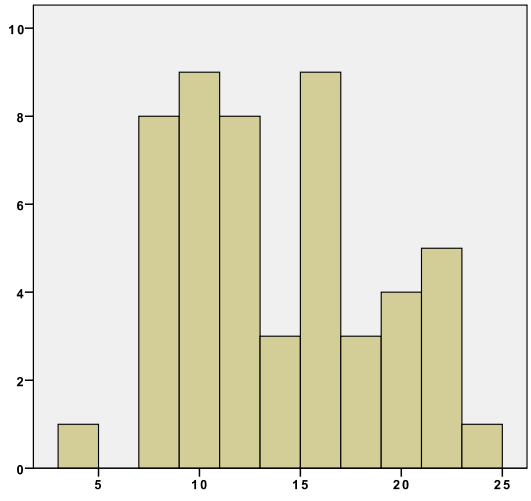

Figure 40. Frequency histogram of Bylsma Q-scores for the Wilson Normal group (N=51). Vertical axis shows frequencies out of 10 and horizontal axis shows scores

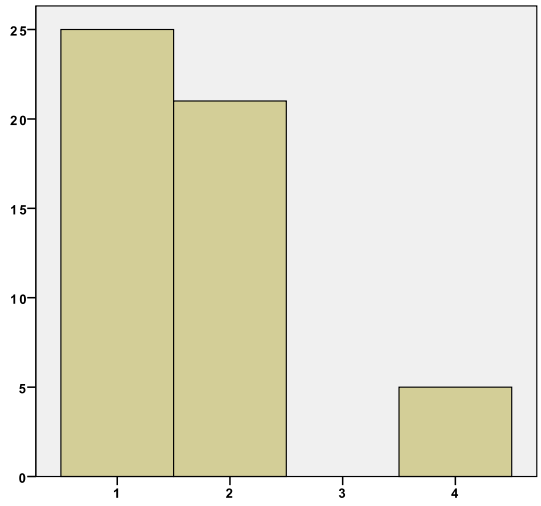

Figure 41. Frequency histogram of Osterrieth scores for the Wilson Normal group (N=51). Vertical axis shows frequencies out of 25 and horizontal axis shows scores.

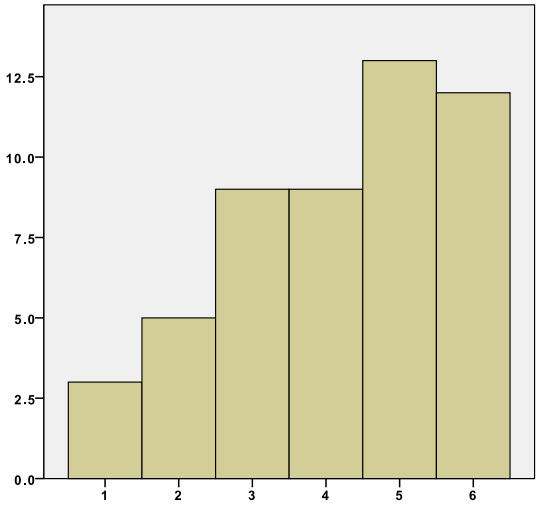

Figure 42. Frequency histogram of Savage Total scores for the Wilson Normal group (N=51). Vertical axis shows frequencies out of 15 and horizontal axis shows scores.

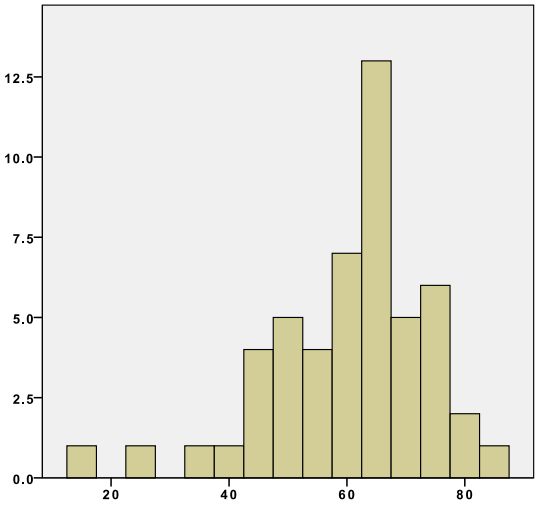

Figure 43. Frequency histogram of Shorr scores for the Wilson Normal group (N=51). Vertical axis showsfrequencies out of 15 and horizontal axis shows scores.

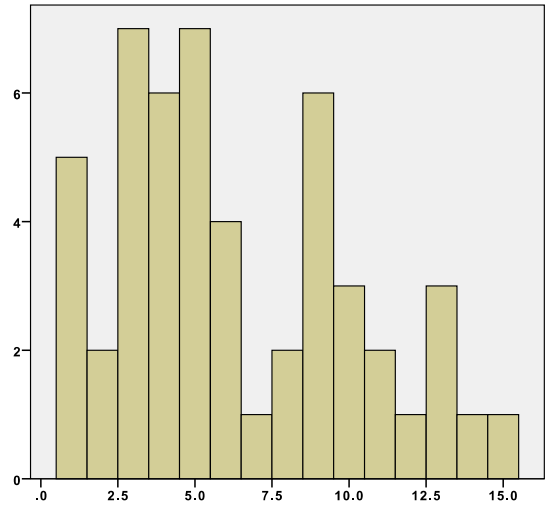

Figure 44. Frequency histogram of Visser scores for the Wilson Normal group (N=51). Vertical axis shows showsfrequencies out of 8 and horizontal axis shows scores.

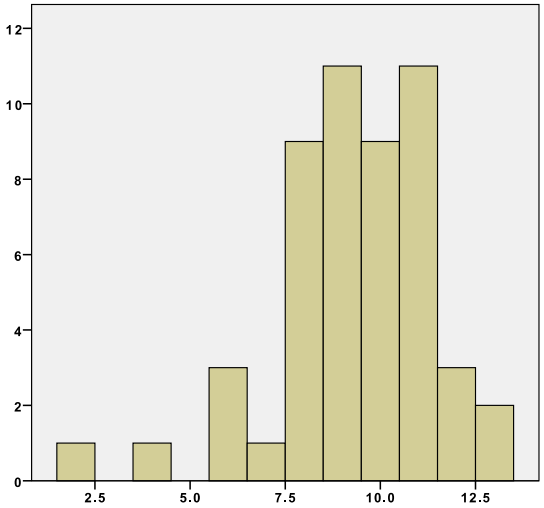

Figure 45. Frequency histogram of Waber scores for the Wilson Normal group (N =51). Vertical axis shows frequencies out of 8 and horizontal axis shows scores.

Appendix 4: Visser (1973) Scoring

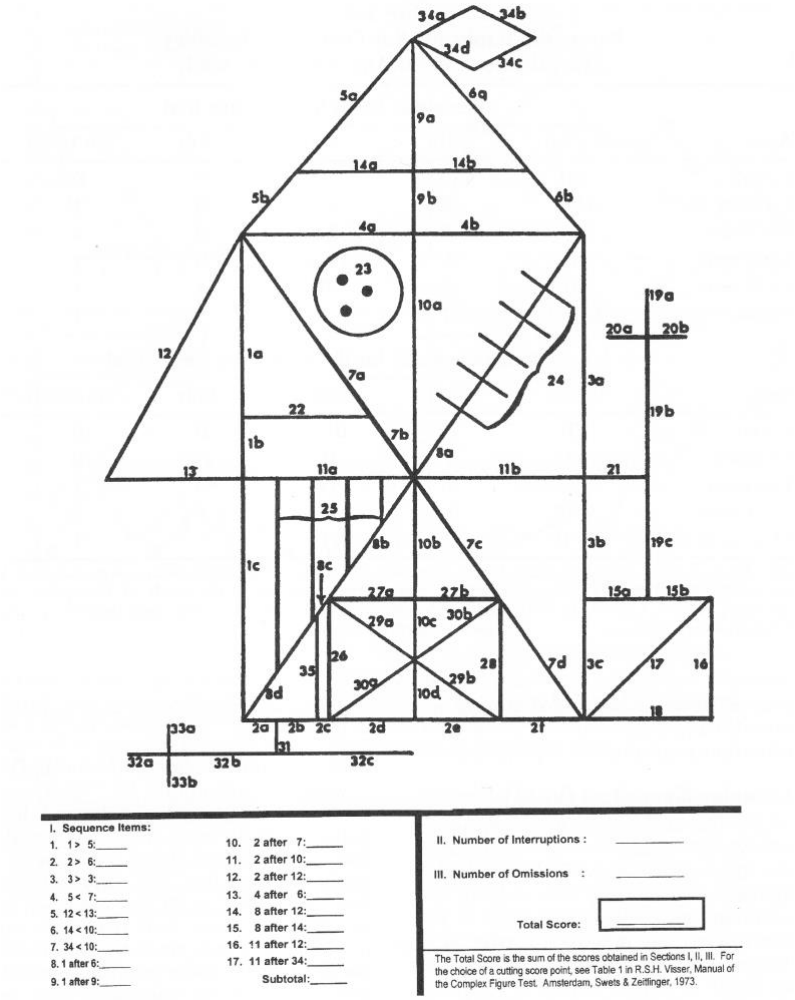

Figure 46. Scoring form for Visser Complex Figure Test. Higher scores denote poorer performances. From Knight, J. A. (2003). ROCF administration procedures and scoring systems. In *The Handbook Of Rey-Osterrieth Complex Figure Usage: Clinical and Research Applications.* Knight, J. A. (Ed.). Lutz, Fl: Psychological Assessment Resources.

Appendix 5: Binder (1982) Scoring

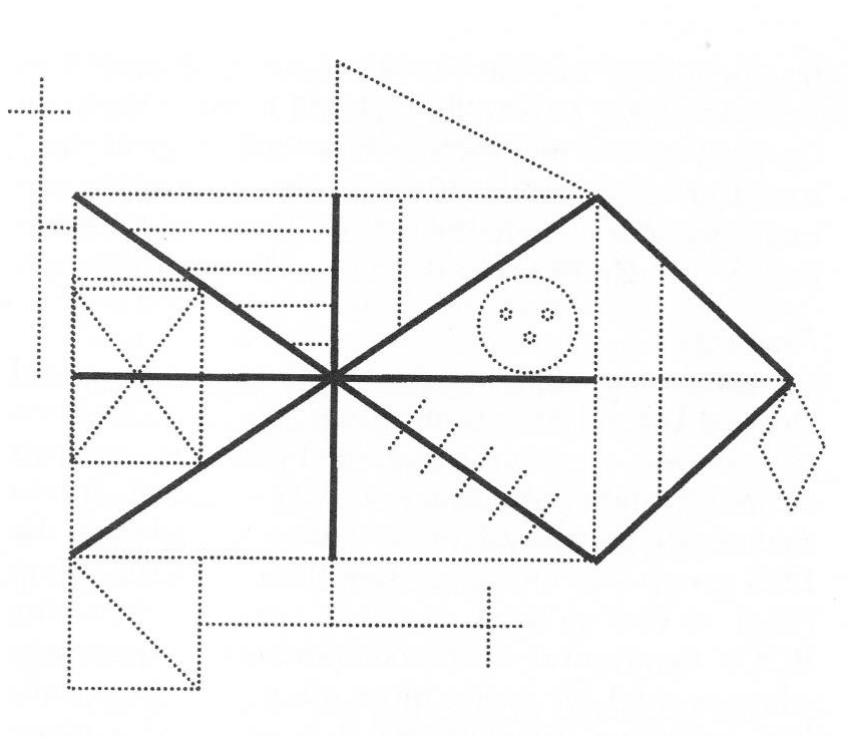

Figure 47. Elements comprising Binder's (1982) Configural Score, for a Total Score of 0-5. From Knight, J. A. (2003). ROCF administration procedures and scoring systems. In *The Handbook Of Rey-Osterrieth Complex Figure Usage: Clinical and Research Applications.* Knight, J. A. (Ed.). Lutz, Fl: Psychological Assessment Resources.

Appendix 6: Bennett-Levy (1984) Scoring – Continuation

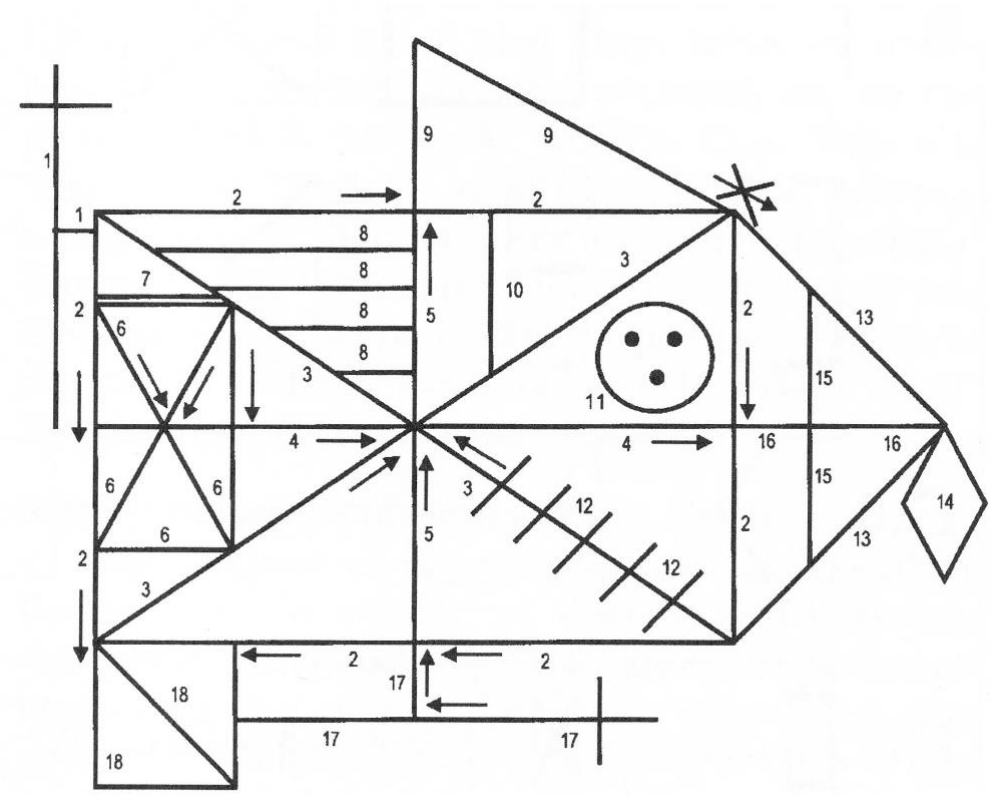

Figure 48. Points of good continuation (denoted by arrows) comprising Bennett-Levy's (1984) Continuation Score. Good continuation points are credited if lines are drawn continuously either in the direction of the arrows, or in the reverse direction. The cross arrow represents a point of poor continuation. Total Scores range from 0-18. From Bennett-Levy, J. (1984). Determinants of performance on the Rey-Osterrieth Complex Figure Test: An analysis, and a new technique for single-case assessment. *British Journal of Clinical Psychology*, 23, 109-119.

Appendix 7: Bennett-Levy (1984) Scoring – Symmetry

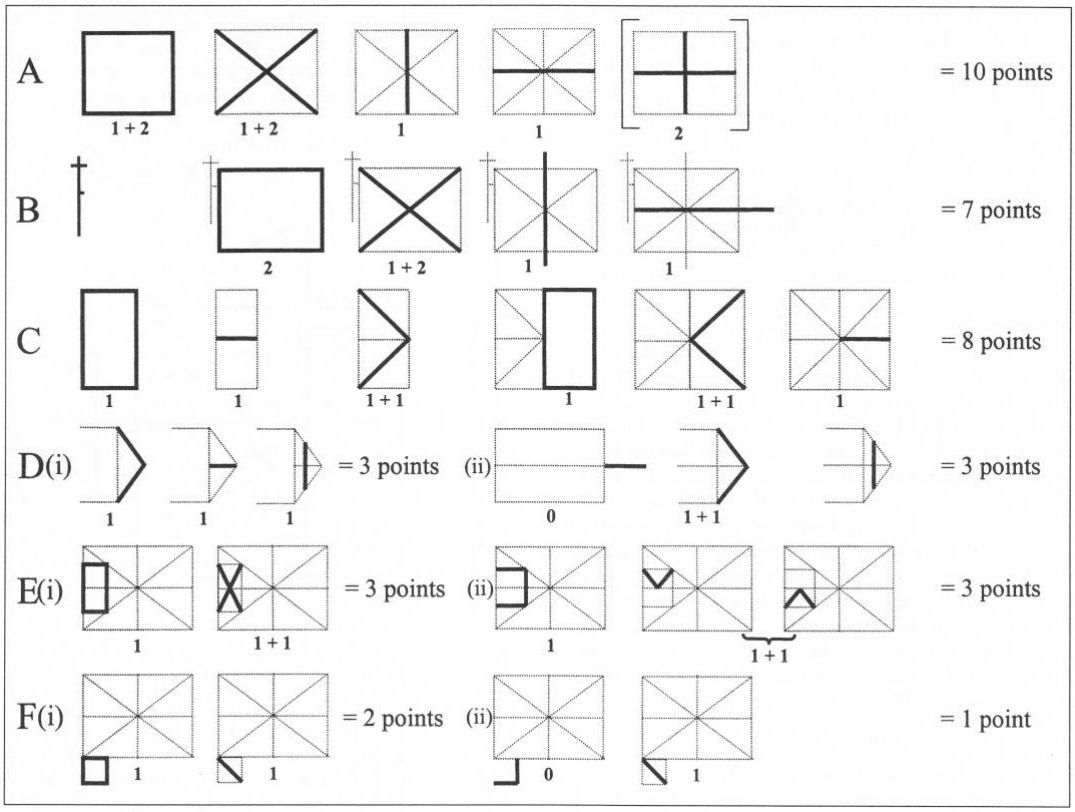

Figure 49. Representations of copying strategies and symmetry points comprising Bennett-Levy's (1984) Symmetry Score. Dotted lines indicate previously drawn elements of the ROCF. Solid lines indicate currently drawn elements. Total Scores range from 0-18. From Bennett-Levy, J. (1984). Determinants of performance on the Rey-Osterrieth Complex Figure Test: An analysis, and a new technique for single-case assessment. *British Journal of Clinical Psychology,* 23, 109-119.

Appendix 8: Shorr et al (1992) Scoring

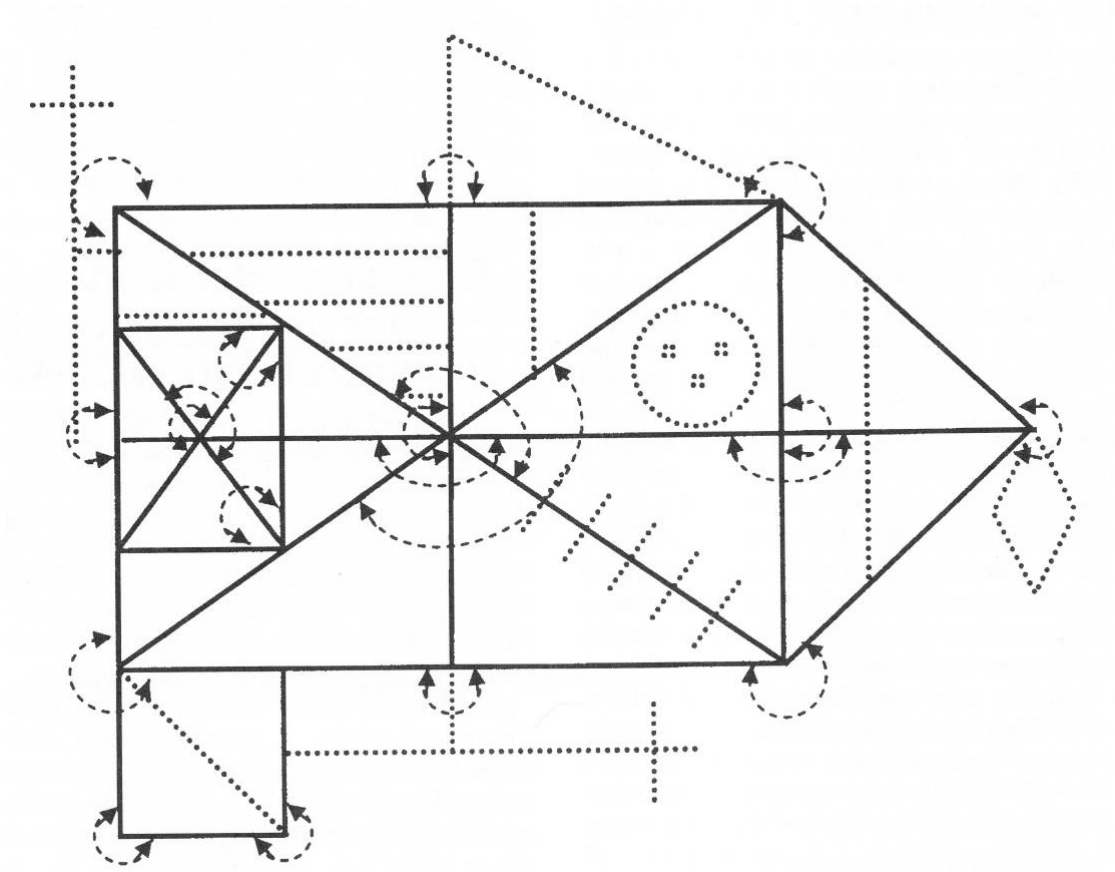

Figure 50. Junctures comprising Shorr et al's (1992) Perceptual Cluster Score are represented by curved dashed lines. Points are credited for each unbroken juncture, for Total Score of 0-20. Perceptual Cluster Ratio derived by dividing Perceptual Cluster Score by number of completed junctures and multiplying this figure by 100. From Shorr, J. S., Delis, D. C., & Massman, P. J. (1992). Memory for the Rey-Osterrieth Figure: Perceptual clustering, encoding, and storage. *Neuropsychology*, 6, 43-50.

Appendix 9: Bylsma (2008) Scoring

The Q-Score:

Rey-Osterrieth CFT Copy Trial Qualitative Scoring

Name/ID: _____ ***Date:*** _____ ***Age:*** _____ ***Sex:*** _____

Education (ys): _____ ***Diagnosis:*** _____

Handedness: **R L** ***Using:*** **R L** ***Scored by:*** _____

Unit	*Line Numbers*	*Unit Score* Y N	*Order*
Base Rectangle	_____	3 0	0 3 *ONLY If lines 1, 2, 3, & 4*
Diagonals	_____	2 0	0 2 *ONLY If lines 5 & 6, 6 & 7, or 7 & 8*
Bisectors	_____	2 0	0 2 *ONLY If lines 5 & 6, 6 & 7, or 7 & 8*
Box with Diagonals	_____	1 0	
Left Cross	_____	1 0	
4 Horizontal Lines	_____	1 0	
Square on Bottom	_____	1 0	
Triangle at Right	_____	1 0	
Top Triangle	_____	1 0	
Circle with 3 Dots	_____	1 0	
5 Crosshatch Lines	_____	1 0	
Diamond	_____	1 0	
Bottom Cross	_____	1 0	

_____ + _____ = _____

Units + *Order* = ***Q-Score***

Completion Time: ___ *min* ___ *sec* = _____ *sec*

Figure 51. Elements and score breakdown for Bylsma et al's (1995) Q-Score, for a Total Score of 0-24. From Bylsma, F. W. (2008). *The Q-Score: A brief reliable method for coding how subjects copy the Rey-Osterrieth Complex Figure.* Chicago, IL: Neuropsychological Services.

Appendix 10: Savage et al (1999) Scoring

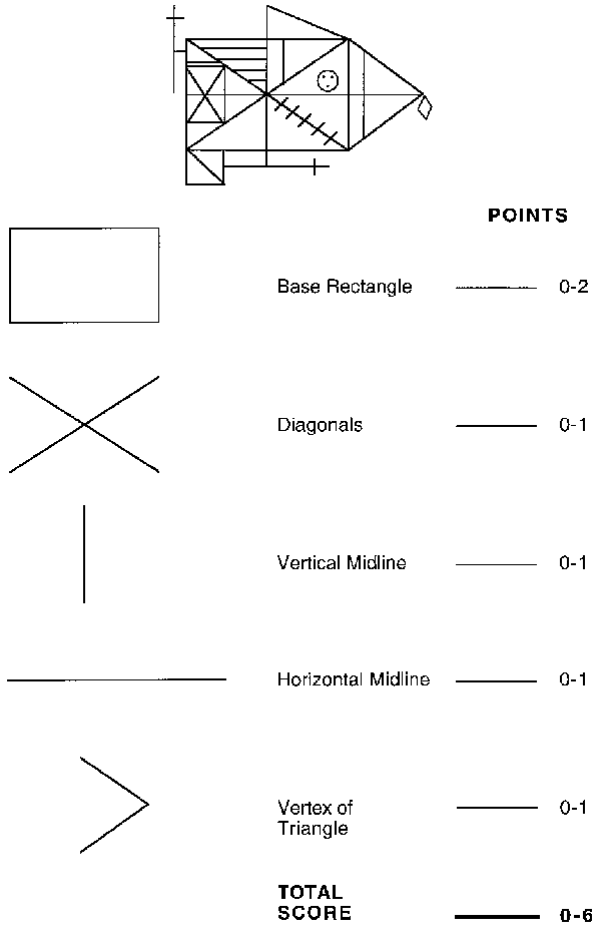

Figure 52. Savage et al (1999) scoring system. Credit is given if each element is drawn as an unfragmented unit, for a Total Score of 0-6. From Savage, C. R., Baer, L., Keuthen, N. J., Brown, H. D., Rauch, S. L., & Jenike, M. A. (1999). Organisational strategies mediate nonverbal memory impairment in obsessive-compulsive disorder. *Biological Psychiatry,* 45, 905-916.

Appendix 11: Bernstein & Waber (1996) Scoring

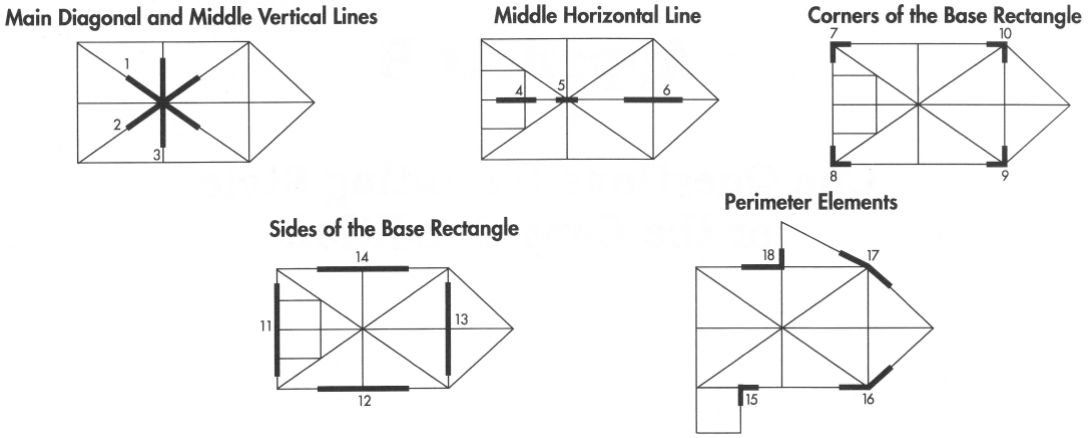

Figure 53. Criterial features comprising Style Score in Waber & Holmes' (1985) scoring system, for a Total Score of 0-18. From Bernstein, J. H., & Waber, D. P. (1996). *Developmental Scoring System for the Rey-Osterrieth Complex Figure: Professional Manual.* Lutz FL: Psychological Assessment Resources.