THE BENDER GESTALT TEST: AN INVESTIGATION INTO PROBLEMS CONCERNING ADMINISTRATION AND SCORING AND ITS APPLICATION TO LOW-EDUCATED ADULTS

by

KATE DYALL

Submitted in fulfilment of the requirements for the degree of

MASTER OF ARTS

in the Department of Psychology
Rhodes University
Grahamstown

1995

DECLARATION

I declare that *The Bender Gestalt Test: An investigation into problems concerning administration and scoring and its application to low-educated adults* is my original work and that all the sources to which I have referred or from which I have quoted have been indicated and acknowledged by means of complete references.

KATE DYALL

ACKNOWLEDGEMENTS

I am grateful to the Human Sciences Research Council for their financial assistance. However, the opinions expressed and conclusions reached are my own, and should not be regarded as a reflection of those of the Council.

My most sincere thanks and acknowledgements go to the following people and organizations:

Ann Jordan for beginning supervision;

Kevin Kelly for taking over supervision of this thesis at a late stage with short notice, for the time he spent familiarizing himself with this subject, his interest, encouragement, humour and pertinent criticism.

John Wiley and Sons who granted permission to include in this thesis Lacks's scoring system from Lacks, P. *Bender Gestalt Screening for Brain Dysfunction Copyright* © (1984 by John Wiley & Sons Inc.);

The Grahamstown and Rini municipalities (now amalgamated into the Grahamstown Transitional Local Council) and particularly the Town Clerk, Steven Cridland, for his cooperation and support in allowing testing, as well as to his staff members;

Enoch Witbooi, leader of the National Union of Municipal Workers, for his time, energy and enthusiasm in supporting this research and for his translation skills, and to the Shop Stewards;

The managers and owners of other organizations in Grahamstown which allowed me to test during working hours, and the Shop Stewards who facilitated finding subjects, particularly Moses Diniso;

My research assistant, Henry Sebata;

Jean Wright, for her most valuable input in proofreading early drafts, and for her support; Sarah Radloff for her dedicated statistical backup; Dinti Mather, who thrashed out the mathematical concept of the conversion of the raw data with me; Darko Dizdar of the Mathematics Department, who assisted with the mathematical principles involved; the Rhodes Library staff for their unwavering assistance and willingness to help; my mother, Julie, for her valuable help in composing tables and in the final editing of this thesis, John Topping for the final proof reading, and close friends who have provided both technical support and constant encouragement.

TABLE OF CONTENTS

PAGE

PAGE

PAGE

PAGE

PAGE

PAGE

PAGE

PAGE

PAGE

LIST OF TABLES

PAGE

PAGE

LIST OF FIGURES

ABSTRACT

The study investigates the use of the Bender Gestalt Test (BGT) amongst low-educated adults. Three versions of the BGT are used in this study; the original 'copy' version as well as the 'immediate' and 'delayed' recall versions. This is done so as to expand the ability of the BGT to identify neurological impairment and to differentiate between this and functional impairment. A literature review explores the problems of standardization in the administration, scoring and application of all three versions of the test. Suggestions are made to correct the problems identified and a novel system of scoring the recall versions are proposed, which allows for the comparison of results of the three versions of the test and which is based on Lacks's (1984) and Weiss's (1970) systems. Administration procedures were also developed to suit the context of the study. The copy, immediate and delayed versions of the BGT were administered to a group of 184 low-educated adults. Statistical analyses revealed significant education effects for the sample tested with regards to both test scores and performance time. The finding of an education effect for performance time is discussed at length, as some literature regards excessive time as a neurological indicator. An anomaly for the group with no education was found to exist, with the scores of these subjects not significantly different from those with 4-6 years of education. Possible reasons for this were explored. In addition, the findings of this research revealed a plateau effect with those having less than 6 years of education scoring substantially lower than those with 7 years and more. The scores of adults with 7 and more years of education level out with no significant differences between educational levels. This appears to suggest that education effects rather than the developmental maturity level proposed by Koppitz, are involved. In addition, the scores of low-educated adults on the expanded Bender Gestalt Test were significantly lower than those of children with similar educational levels, in other studies. These findings and possible explanations are discussed. The study concludes by suggesting new research areas and emphasizing the urgent need for separate normative data on the expanded BGT for low-educated adults, and the establishment of appropriate 'cut-off' points.

PREFACE

This research is presented in four parts which are detailed below.

Part One: Literature Review and Implications for this Study examines the available literature regarding the copy version of the Bender Gestalt Test, the role of memory in brain damage and research done on the Bender Gestalt Recall Tests.

Part Two: Methodological Rationale for the expanded BGT takes a more detailed look at specific aspects of the literature in order to resolve problematic issues regarding the scoring and the administration on both the Bender Gestalt Test (copy version) and the Bender Gestalt Recall Tests. The criteria for the selection of subjects are also considered.

Part Three: Methodological Procedures concerns the actual methodology used in this research. A scoring system for the Bender Gestalt Recall Tests is provided and then the research method is described.

Part Four: Presentation and Evaluation of Research Findings, concludes the research with a presentation of the results of this research, a discussion thereof, and the conclusions reached.

The following abbreviations are used for ease of reference: The Bender Gestalt Test is abbreviated to the BGT (see 2.1.3: p.8), while the copy version is referred to as the BGT-*c*. The Bender Gestalt Recall Tests are abbreviated to the BGRTs in their entirety, while the BGRT-*i* refers to an immediate recall test and the BGRT-*d* to a delayed recall test, preceded by a 10 minute delay. The three tests combined are referred to as the expanded BGT. Finally, the Bender Gestalt designs are abbreviated to the BG designs. For brevity, the term 'low-educated adults' is used to refer to adults who have less than 9 years of education.

This thesis contains an extra section, Part Two: Methodological Rationale of the expanded BGT. The reason for this is that new questions emerged as the research progressed. As literature was reviewed it became evident that a number of unresolved issues on the use of the BGT existed, which included the broad variety of scoring systems and administration procedures. Dana, Field and Bolton (1983, p.81) state that "there is no single test entity properly called the BGVMT since the designs themselves vary, standard and consensual procedure is lacking, and many different forms are used". To have arbitrarily decided upon a scoring system and an administration method without considering all the available options and the appropriateness of these options, would have been to perpetuate the confusion of research which surrounds the BGT. Instead it was decided to investigate the scoring and administration issues in a comprehensive manner, which resulted in the addition of a second literature review concerning the rationale of the methodology finally used in this study.

Such a problem encountered during research is not unique and can be part of the research process itself. In a discussion of developing a method for participatory research of a hermeneutic nature, Kelly and Van Vlaenderen (1994, p.11) stated that "the conclusion and literature reviews continued to influence each other until the point of conclusion". This was true of the current research. The formulation of research questions is inherent to the progression of the research, with the aim that more thorough questions are asked and an improved manner of explicating the subject under investigation is attained (Kelly & Van Vlaenderen, 1994). These authors add that a vital tenet of this type of research method is that "the data does not speak without being asked questions and the process of asking more perspicacious, better defined questions, leads to knowing a phenomenon better ..." (Kelly & Van Vlaenderen, 1994, p.11). An increased knowledge of the subject under investigation once again creates more questions. Thus, while certain issues were clarified and problems resolved in the process of this research, more questions arose from the increased understanding of the subject. In keeping with the opinions of Kelly and Van Vlaenderen (1994) above, these research findings do not solve all the questions surrounding the use of the expanded BGT, but rather add new areas which require further investigation.

Figure 1 (p.xix) clearly indicates the structure of this research as outlined above.

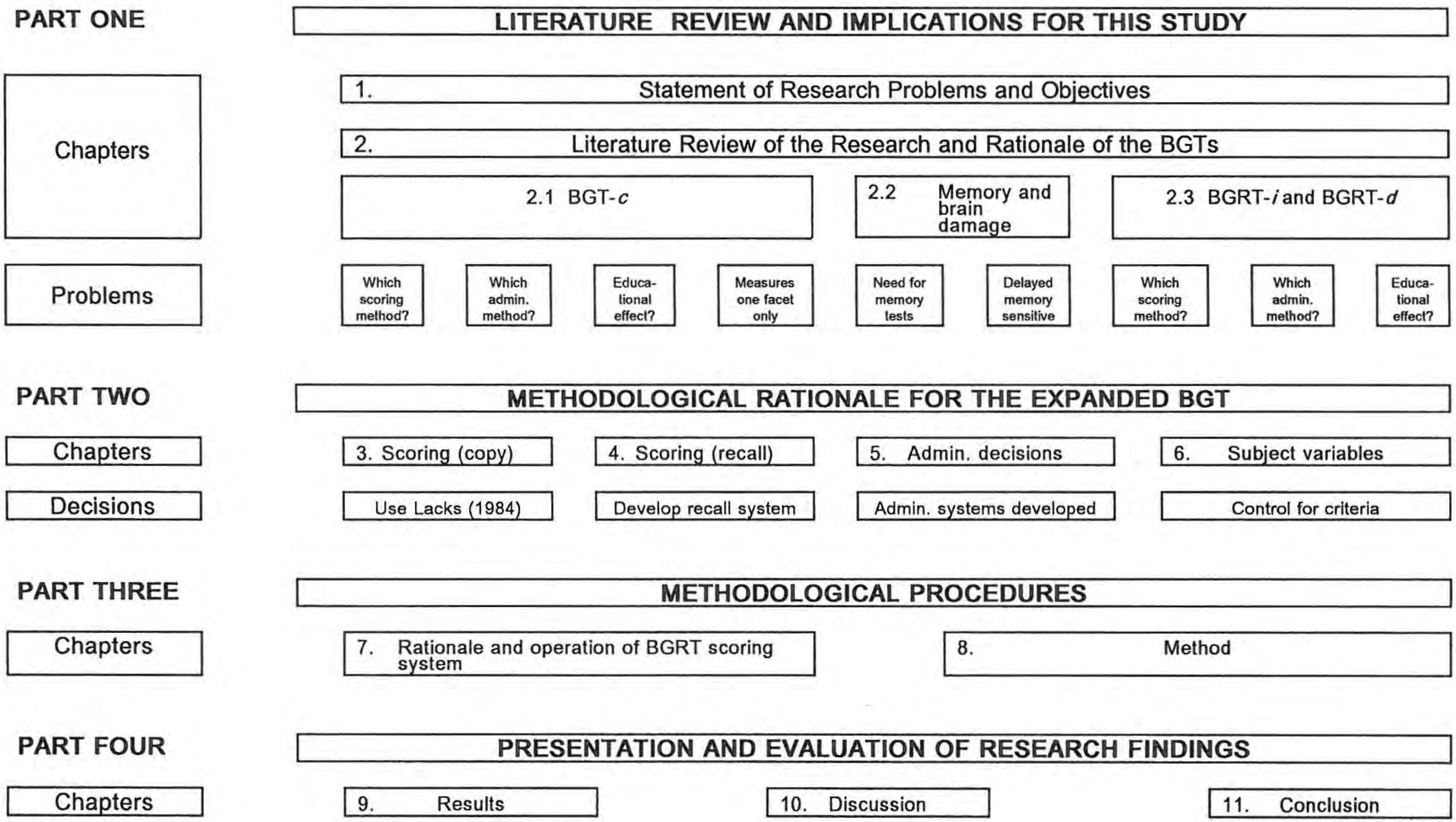


FIGURE 1: A DIAGRAMATIC REPRESENTATION OF THE STRUCTURE OF THIS THESIS

PART ONE: LITERATURE REVIEW AND IMPLICATIONS FOR THIS STUDY

Part One begins with Chapter 1: A Statement of the Research Problems and Goals. Chapter 2 introduces the literature on the BGT-*c*. This test is then critically evaluated. Next the key role played by memory impairment in brain damage is reviewed, followed by literature concerning the BGRTs. Chapter 2 concludes with a proposal suggesting the need for research such as is undertaken in this thesis. This ends Part One.

CHAPTER 1: STATEMENT OF RESEARCH PROBLEMS AND OBJECTIVES

The BGT-*c* is a widely available test which is rapid to administer. Comprehensive literature reviews and surveys indicate that the BGT-*c* is one of the most useful and extensively used neuropsychological tests in the screening for and identification of neurological impairment (Dana *et al.*, 1983; Hain, 1964; Lubin, Larsen & Matarazzo, 1984; Lezak 1983; Schulberg & Tolor, 1961). Research shows that the BGT-*c* is a good measure for differentiating between organic and functional impairment in both children and adults (Allen, Adamo, Alker & Levine, 1971; Butler, Coursey & Gatz, 1976; Hutton, 1966; Lacks, 1984; Lubin & Sands, 1992; Maloney & Ward, 1970; Mosher & Smith, 1965; Wagner & Murray, 1969). This is important, as presenting symptoms of organic and functional impairment can appear very similar (Coleman, Butcher & Carson, 1984; Kaplan & Sadock, 1991). This can be problematic to diagnoses and consequently to appropriate treatment.

However, despite its assets, the use of the BGT-*c* is hampered by a number of problems. Firstly, although cerebral damage has a multifaceted presentation, memory dysfunction is often the first presenting symptom of such impairment (e.g. Lezak, 1983). Additionally, memory impairment is usually one of the behavioural manifestations of brain damage, regardless of the aetiology (Lezak, 1983; Lishman, 1978; Walsh, 1991; Williams, 1991). The assessment of brain damage should therefore include the degree of immediate memory retention (e.g. Lezak, 1983; Lishman, 1978; Walsh, 1991). Delayed memory tests can assist in the differentiation of specific features of cerebral damage as well as the location thereof (e.g. Lezak, 1983). However, memory is not measured by the BGT-*c*. Thus, expanding the BGT-*c* into a memory test by adding a BGRT-*i* and a BGRT-*d* will arguably enhance its diagnostic value. However, for the expanded BGT to be really useful in daily assessment, it must be practical and economical to administer and rapid to score.

Despite the popularity and usefulness of the BGT-*c*, it is marred by many inconsistencies (Dana *et al.*, 1983; Eno & Deichmann, 1980). The BGT-*c* has been criticized for its lack of standardization (e.g. Billingslea, 1963; Dana *et al.*, 1983; Popplestone, 1956) and many studies on the BGT-*c* exist which cannot be replicated. Further, one of the most profound problems is the proliferation of methods used to evaluate this test (e.g. Bender, 1948; Hain, 1964; Koppitz, 1975; Lacks, 1984;, Pascal & Suttell, 1951; Pauker, 1976) and certain deficiencies in most of these methods. Evaluation procedures range from subjective judgement (Bender, 1938; Dana *et al.*, 1983; Mosher & Smith, 1965; Wagner & Murray, 1969), through a proliferation of scoring systems with specific evaluating criteria (e.g. Hain, 1964; Koppitz, 1975; Lacks, 1984; Pascal & Suttell, 1951; Pauker, 1976), to the extreme of the 312 weighted indices of Gobetz (1958, in Dana *et al.*, 1983). Most of these evaluation methods are defective in some way (e.g. Butler *et al.*, 1976; Dana *et al.*, 1983; Hain, 1964; Koppitz, 1975; Lacks, 1984; Lezak, 1983; Mosher & Smith, 1965; Pascal & Suttell, 1951; Wagner & Murray, 1969).

Lezak (1983, p.388) stated that "the profusion of scoring possibilities has resulted in many attempts to develop a workable system to obtain scores for diagnostic purposes". Despite the number of systems, until recently such a "workable system" (Lezak, 1983, p.388) appeared to be nonexistent. Clearly, different methods of evaluation make comparison of research findings problematic. Further, since three protocols per person need to be evaluated when using the expanded BGT, rapidity of scoring becomes paramount. The use of comparable criteria for the copy and recall phases expedites marking, as only a single set of scoring concepts needs to be mastered. Also, uniformity of scoring across the copy and recall versions is important as it allows for comparisons between the two test modes, without the confounding factor of unmatched design evaluation.

A literature review isolated Lacks's[1] (1984) adaptation of the Hutt-Briskin scoring system as highly appropriate for use in the current research. Lacks's system (1984) is an

---

1. According to Gowens (1977) this is the correct way to refer to a possessive pronoun that ends with an 's'.

operationalization of Hutt and Briskin's (1960) 12 indicators of brain damage. It adheres to the concept of the gestalt, which forms the fundamental rationale of the BGT, and is further discussed later in this research. Also, it is designed to detect brain damage in adults, has high reliability and validity and a high accuracy rate for novice scorers. Most important, with Lacks's system (1984) protocols can be scored rapidly. Prior research (Dyall, 1993) combined Lacks's (1984) scoring system for adults on the BGT-*c* and Weiss's (1970) BGRT scoring system for children, to produce a system for the evaluation of the two BGRTs. Thus, a system was devised which circumvented some of the difficulties outlined above, in that it ensured rapid scoring as well as comparability between the BGT-*c* and the BGRTs. This scoring system was used effectively to compare performances of children, aged between 7,5 and 16 years old, on the expanded BGT (Dyall, 1993).

Research reveals that educational levels influence scores significantly on neuropsychological measures (Cronbach, 1991; Ernst, 1987; Finlayson, Johnson & Reitan, 1977, Kennedy, 1981; Leckliter & Matarazzo, 1989; Prigatano & Parsons, 1976). Education is also positively related to performance on neuropsychological tests even where functions are apparently unrelated to academic performance, as with immediate memory tests (Lezak, 1983). On the BGT, education effects have been found (Brilliant & Gynther, 1963; Carlson, 1966, Koppitz, 1975; Lacks, 1984; Smith & Winnick, 1979). In South Africa, clinicians often work with adults who have little or no education, for example, an incomplete high school education. South Africa has a high illiteracy rate among 'other-than-whites', for whom schooling has been neither compulsory nor available. In the Eastern Cape, for example, an estimated 30% of admissions at Fort England Mental Hospital, which currently services the region, have little or no education (S. O'Donoghue, personal communication, 6 May 1994). This estimation is concomitant with worldwide adult illiteracy, estimated to be 900 million, which constitutes approximately one-third of the world's adult population (Steinberg, 1990). The BGT is often used for preliminary evaluation on admission and norms or organic 'cut-off' points for well educated adults are available on a variety of scoring systems (Hain, 1964; Lacks, 1984; Pascal & Suttell, 1951; Pauker, 1976). However, to the author's knowledge, no norms for adults which have been specifically stratified for very low levels of education exist in published literature. Norms for the BGRT-*i* for the population in question are also not available. A BGRT-*d*, after a 10 minute interval, is a feature introduced by Dyall (1993). For this version of the test, no norms are available for adults on any educational level.

Clearly, normative data for a population with a completed high school and/or university education are inappropriate to adults who have less than 9 years of education. The establishment of preliminary normative data for such adults in a South African population is necessary and urgently needed in order to improve the accuracy of diagnostic assessments. The Lacks (1984) scoring system has been effectively used on children with as little as 2,5 years education (Dyall, 1993) and as such, is likely to be applicable to this sample of adults.

Bearing the above in mind, the following research study sets out to achieve six objectives corresponding to each of the problems outlined above namely: (i) to enhance the diagnostic value of the BGT by extending it with a BGRT-*i* and a BGRT-*d*; (ii) to establish standardized administration procedures for these three tests which are economical and appropriate to low-educated adults; (iii) to explore the use of a scoring system for the BGRTs which is compatible with the scoring system used on the BGRT-*c* and rapid to score; (iv) to operationalize this method in order to make it accessible to clinicians; (v) to investigate the effects of education on the expanded BGT with low-educated adults; (vi) to establish age-stratified preliminary normative data for a non-clinical population from a sample of South African adults with fewer than 9 years of education, on the BGT-*c* scored according to the Lacks' system, and on the two BGRTs evaluated by a scoring system derived from a combination of Lacks's (1984) BGT-*c* and Weiss's (1970) BGRT scoring systems.

CHAPTER 2: LITERATURE REVIEW OF THE RESEARCH AND RATIONALE OF THE BGT-*c* and BGRTs

A review of the BGT literature and other literature relevant to the expansion of the BGT-*c* with two BGRTs is presented in five main sections. First, literature concerning the BGT-*c* is reviewed. Secondly, this literature is critically evaluated. Thirdly, the role of memory in brain damage is reviewed. Fourthly, research on the BGRTs are presented. The fifth section draws the contents of this review together and concludes Part One. The deficits which have been noted result in the proposal of the purpose of this study. Finally problems which still need to be addressed are outlined.

2.1 A literature review of the BGT-*c*

This section reviews literature which considers the BGT-*c*, beginning with its origin and history. This is followed by a discussion which situates the test within the context of gestalt psychology, the discipline which forms the underlying rationale of the test. Next is a brief consideration of terminology. The popularity and uses of the BGT-*c* are then appraised emphasizing this test's proficiency in the identification of cerebral dysfunction and its value in discriminating between functional and organic impairment. Following this, the use of the BGT-*c* as a diagnostic screening test is discussed and the different versions of the BGT-*c* are described. Next the functions measured by the BGT-*c* are considered. This section concludes a detailed review of BGT-*c*'s ability to identify neurological impairment and to differentiate between organic and functional impairment.

2.1.1 Origin and history of the BGT

The BGT-*c* was developed by Laurette Bender and was first published in 1938. She selected nine of the forms originally used by Wertheimer in his renowned 1923 paper (Bender, 1938) which documents his experiments with visual gestalten. Billingslea (1963) notes that Bender first began using Wertheimer's designs prior to 1932. The original version of Bender's (1938) test was a diagnostic, non-scored test, administered by requesting subjects to copy the nine selected designs, which subsequently became the BG designs or stimuli. Bender (1938) used these designs in the course of her investigations of intelligence in children and in functionally and neurologically impaired individuals. She evaluated the protocols subjectively for indications of functional or organic impairment. Although many variations of the BGT have been introduced since its inception, this version is still the most frequently used as studies reviewed in this section will show. The tenets of the BGT are derived from the discipline of psychology from which its name was derived, namely, gestalt psychology.

2.1.2 Situating the rationale of the BGT within the discipline of gestalt psychology

Gestalt psychology is founded on the philosophy of Wertheimer, Köhler and Koffka. It emanated from physics and mathematics and is grounded in these fields (Köhler, 1947). Gestalt psychology emerged during the 1920s in reaction to the behaviourist movement which regarded direct experience as subjective and therefore not accessible to direct observation. Because of this, the behaviourists regarded subjective experience as irrelevant to scientific research (Köhler, 1947).

In contrast to the behaviourist approach, gestalt psychology maintains that subjective experience is a vital component of neurological functioning (Köhler, 1947). From the time of its inception, this new paradigm in psychology asserted that inferences based on observation were necessary since the functioning of the central nervous system cannot be directly observed (Köhler, 1947). Gestalt adherents therefore made deductive inferences about the functioning of the central nervous system from the outcome of subjective perceptual experiences (Köhler, 1920, 1947; Petermann, 1932; Wertheimer, 1922, 1923, 1925). Rejecting the mechanistic theories of perception prevalent at the time (Köhler, 1947), gestalt psychology advanced the theory that perception cannot be regarded as merely the combination of a number of events (Schilder, 1938). This became the founding principle of gestalt theory which states that the whole (i.e. the gestalt) is greater than the sum of the parts. 'Gestalt' is the term used for the result of the total perceptual process (Köhler, 1920). In other words, the total perceptual process is always more meaningful than the mere combination of the individual components which contribute to this process (Köhler, 1920, 1947; Petermann, 1932; Wertheimer, 1922, 1923, 1925). Köhler (1920, p.17) describes the process by which gestalten develop as follows:

> When spatial, visual, auditory and intellectual processes are such as to display properties other than could be derived from the parts in summation,

> they may be regarded as unities illustrating what we mean by the word 'Gestalten'.

The rationale underlying this hypothesis concerns the process by which the gestalten are organized. Gestalten are organized according to specific intrinsic principles (Köhler, 1947; Wertheimer, 1922, 1923) based on tendencies inherent in the human perceptual process. Therefore, it can be said that perception comprises the overall combination of stimuli combined with the total operations which are executed. Perception is thus the response of an individual's entire organism to the full set of circumstances present at the time. Schilder (1938, p.vii) describes perception as the organisms's reaction to "constellations of stimuli by a total process" rather than the combination of single events. Bender (1938) states that information which is perceived, is processed and integrated according to patterns, by the central nervous system. The result is that the final patterns formed, also called gestalten, are circumscribed or determined by the central nervous system.

Gestalt configurations, such as those used by Bender (1938) in her test, are also perceived in terms of pattern. The final gestalten or patterns produced by an individual copying the designs is the outcome of the "interplay of the dynamic organization of the total organism with the [total] situation" (Schilder, 1938, p.vii). As perception is influenced by experience and adaptation, these two factors, experience and adaptation, influence the way in which gestalten are completed and reorganized by the central nervous system (Bender 1938).

The BGT is based on this tenet of the organization and integration of patterns. Since the gestalt function is one of holistic integration, Bender (1938) correctly anticipated that gestalt configurations would vacillate according to both the maturational stage of an individual as well as to organic and functional impairment. She found that perceptual gestalten regressed to more primitive stages when disintegrating cerebral lesions existed and returned to the higher integrative responses as the brain recovered. This accords with the tenets of developmental maturity theories. Through extensive research, Bender (1938) established that the higher cerebral centres are likely to be the nuclei of intricate types of organization. In addition, she maintained that lower central nervous system centres such as the spine and peripheral nerves were also involved in perceptual organization. Bender (1938) reasoned that damage in any of these locations would lead to disturbances in the gestalt formation.

A highly significant principle of the gestalt is that it operates according to the tenets of "primacy of the whole over the part" (Wertheimer, in Petermann, 1932, p.47). In other words, the overall pattern rather than the individual components of a pattern are perceived. Where the organization of the perceptual process is defective, this primacy can be compromised. One of the fundamental criteria emphasized in some of the scoring systems of the BGT is the *breaking up* of the gestalt (e.g. Lacks, 1984). For example, primacy is compromised when a configuration is perceived and copied as individual components, with disregard to the overall design. In such cases the individual parts take precedence over the total pattern, which is derived from the combination of these components plus the individual's experience and organizational and integrating abilities.

Schilder (1938, p.vii) introduces the BGT as the first "systematic approach" to the problem of perception, adding that "the results obtained in schizophrenia, manic depressive psychosis, aphasias [*sic*] and organic brain diseases get their full meaning when one compares them with the standardized lines of development." Bender (1938, p.3) cogently sums up the significance of the reproductions of the BG designs by stating that "the final product is a visual motor pattern which reveals modifications in the original pattern by the integrating mechanism of the individual who has experienced it."

The principles of the BGT are embedded in gestalt psychology; a fact which will be shown to be very significant in the course of this research. For example, one of the major problems with a number of systems used to score the BGT is their failure to consider the underlying tenets of gestalt psychology. Some systems are in fact contrary to the concept of the gestalt, which has been outlined above.

2.1.3 The variety of names used to refer to the BGT

Bender (1938) originally called the test which she developed, "A Visual Motor Gestalt Test". However, Bach (1987, p.109) points out that a confusing variety of names and "curious permutations" thereof have been used in published literature. Often a variety of different nominations appear in a single paper. For example, Lubin, Larsen, Matarazzo & Seever, 1985, p.858) use three terminologies for the test, namely "the Visual-Motor Test (Bender Gestalt)", "the Bender-Gestalt Visual-Motor Test" and the "Bender-Gestalt". Bach (1987) points out that the nomenclature, Bender-Gestalt Visual Motor Test, is inaccurate, as the sequence is confused. Instead it should be called the Bender Visual-Motor Gestalt Test.

A profusion of other name variants have occurred (e.g. Anderson & Rallis, 1981; Arbit & Zager, 1978; Brilliant & Gynther, 1963; Choynowski, 1970; Flick & Duncan, 1973; Friedman, Strochak, Gitlin & Gottsagen, 1967; Frostig, Lefever & Whittlesey, 1961; Obrzut, Taylor & Thweatt, 1972; Sabatino & Ysseldyke, 1972; Smith & Winnick, 1979; Thomas, 1984). Some of the more unusual appellations include the "Bender Visual Motor Gestalt Designs Test" (Arbit & Zager, 1978, p.460), "the Bender Gestalt Test of Visual-Motor Coordination" (Collaer & Evans, 1982, p.642), "the Koppitz Bender Gestalt Test" (Holroyd, 1966b, p.440) and the "Bender-Koppitz Test" (Choynowski, 1970, p.135). Many of the nominations of the BGT omit the word 'gestalt' which denotes the fundamental principle upon which the entire test is based. Instead, researchers refer to the test as 'the Bender' or the 'Bender Test' (Anderson & Rallis, 1981; Brilliant & Gynther, 1963; Choynowski, 1970; Friedman *et al.*, 1967; Koppitz, 1975; Obrzut *et al.*, 1972; Sabatino & Ysseldyke, 1972; Thomas, 1984).

Bach (1987, p.109) states that the omission of the word 'gestalt' from the name of the test is "unjustified and misleading" since the designs originated from Wertheimer's classic study of gestalt principles. He adds that some psychologists believe that the test should have been called the 'Wertheimer-BGT' but that it is now too late to introduce such a change (Bach 1987, p.111). Nevertheless, Bach (1987) argues that it is mandatory to retain the term 'gestalt' in recognition of the origin of the designs and their underlying principles.

Hutt's (1945, in Bach, 1987) unpublished manuscript first added Bender's name to the test and shortened it to the 'Bender-Gestalt Test'. Bach (1987) maintains that this nomination is suitable, as is the shortened form - BGT. In addition, this appears to be one of the most frequently used nominations (e.g. Bender, 1965; Dana *et al.*, 1983; Eno & Deichmann, 1980; Lacks, 1982, 1984; Lacks & Newport, 1980). It was therefore decided to use the nomenclature, 'BGT'.

2.1.4 The popularity, uses and value of the BGT

The BGT-*c* has been an approved and effective clinical assessment tool for over half a century (Dana, *et al.*, 1983; Hain, 1964; Lezak, 1983; Lubin, *et al.*, 1984; Schulberg & Tolor, 1961). It has been described both as "popular and influential" (Dana *et al.*, 1983 p.76).

Surveys done over the years reflect a consistently high usage of the BGT-*c* in assessments and evaluations. Schulberg and Tolor (1961) surveyed 176 clinicians, 94% of whom reported having at least 5 years diagnostic evaluation experience. Of these, 85% stated that they used the BGT-*c* when testing adults and 68% said that they used this test when testing children. Lubin *et al.* (1984) maintained that the BGT-*c* was the third most frequently used test over five decades. Using the same data from the study just mentioned, Lubin *et al.* (1985) calculated the frequency of usage of various tests. They reported that the BGT-*c* was used 80% of the time, in five different types of mental health institutions.

However, Reynolds (1979) points out that the *frequency* with which a neuropsychological test is used is not predominately related to the actual *quality* of the test itself. In a study investigating the quality of psychometric tests, Reynolds (1979, p.326) surveyed 41 psychologists to establish their views of the "overall quality of psychometric refinement" of the 10 most used tests. The BGT-*c* ranked fifth. In addition, the BGT-*c* emerged as one of five tests most recommended by 471 American Psychological Association members for students to learn to administer and score (Wade & Baker, 1977; Wade, Baker, Morton & Baker, 1978). Also regarding the actual value of the BGT-*c*, Lubin and Sands (1992) compiled a bibliography spanning 21 years of studies which endorses the psychometric abilities of the BGT-*c*. They note that "there is a substantial literature on the reliability and validity" of the test (Lubin & Sands, 1992, p.385). Since its inception in 1938, research conducted using the BGT-*c* has investigated a broad variety of conditions in both adults and children which is illustrated by the review below.

The BGT-*c* is a popular measure for assessing visual motor integration and overall development in children (e.g. Frey & Pinelli Jnr, 1991; Isaac, 1973; Kaspar & Lampel, 1972; Koppitz, 1975; Pope & Snyder, 1970; Ryckman, Rentfrow, Fargo & Mc Cartin, 1972; Schachter, Brannigan & Tooke, 1991; Taylor & Thweatt, 1972). In addition, this test also successfully identifies visual perceptual problems in deaf children (Gilbert & Levee, 1967).

It is extensively applied in the field of child education to predict learning disabilities (e.g. Henderson, Butler & Goffeney, 1969; Nielson & Sapp, 1991; Skeen, Strong & Book, 1982; Thweatt, 1963). The BGT-*c* measures the achievement of reading skills and identifies reading disabilities (e.g. Black, 1973; Clark, 1982; Clarke & Leslie, 1971; Keogh, 1965; Koppitz, 1975; Leviton, Kirby, Guild-Wilson & Neff, 1993; Mc Kay & Neale, 1985; Mc Manis, Figley, Richert & Fabre, 1978; Obrzut *et al.*, 1972; Sabatino & Ysseldyke, 1972; Skeen *et al.*, 1982). Further, it predicts arithmetic ability and detects difficulties in this area (Henderson, *et al.* 1969; Koppitz, 1962; Leviton *et al.*, 1993).

Clinicians over a long period have also used this test as a projective technique for personality evaluation in samples of pathological populations as well as with clinically normal individuals (e.g. Bender, 1946; Even, Kipper & Yehuda, 1988; Haynes, 1970; Hutt & Briskin, 1960). Oas (1984) used it as gauge of impulsivity and Yulis (1970) utilized it to investigate motivational drive. Taylor (1965) found it to be a measure of intelligence and adjustment in individuals whose IQ fell into the lower range. Further, the BGT-*c* also appears to be a prognostic indicator of mental illness (Swensen & Pascal, 1951). In addition to the above studies, the BGT-*c* has been cited as a "rapport builder" (Lubin & Sands, 1992, p.385), an indicator of mental age in children (Billingslea, 1963), a measure of ego strength (Field, Bolton & Dana, 1982), and an indicator of specific types of psychopathologies (Dana *et al.*, 1983; Lubin & Sands, 1992).

While noting the broad variety of uses of the BGT-*c*, the focus of this research is its ability to detect brain damage and to differentiate between functional and organic impairment. The BGT-*c* is most extensively used to screen for and to diagnose cerebral dysfunction (Eno & Deichmann, 1980). It is, in fact, regarded as either the most, or one of the most, extensively used neuropsychological assessment tools in the identification of impaired functioning linked to brain damage (Dana, *et al.*, 1983; Hain, 1964; Lezak, 1983; Lubin *et al.*, 1984; Schulberg & Tolor, 1961). Schulberg and Tolor (1961) found that over half the clinicians they surveyed, used the BGT-*c* in at least 75% of their cases in which possible neurological impairment was implicated. Further, according to a survey of major authorities, the BGT-*c*'s ability to diagnose brain damage is unquestioned and forms the only point of consensus regarding the use of the test (Dana *et al.*, 1983).

Many studies conducted on both adults and children support these views (e.g. Friedman *et al.*, 1967; Frostig *et al.*, 1961; Hanvik, 1951; Holroyd, 1966a; Koppitz, 1962; Korman & Blumberg, 1963; Landis, Baxter & Patterson, 1974; Russell, 1976; Smith & Martin, 1967; Wagner & Murray, 1969; Welcher, Wessel, Mellits & Hardy, 1974). It is also able to identify neurological impairment in individuals with a borderline IQ (Johnson, Hellkamp & Lottman, 1971). The BGT-*c* can identify the presence or absence of retardation (e.g. Allen, *et al.*, 1971; Black, 1973; Song & Song, 1969) as well as levels of retardation (Andert, Hustak & Dinning, 1978; Condell, 1963).

More important, the BGT-*c* is able to differentiate between functional and neurological impairment. In adults, the BGT-*c* has differentiated between normal, psychiatric, and brain damaged adults (e.g. Goldberg, 1959; Hain, 1964; Jernigan, 1967; Lacks, Colbert, Harrow & Levine, 1970;) as well as mixed groups of organic and psychiatric patients (e.g. Brilliant & Gynther, 1963; Canter, 1966; Griffith & Taylor, 1960; Lacks & Newport, 1980). This ability to discriminate has also been applied to children. For example, the test differentiated between emotionally and organically disturbed children (Hilgert, 1985; Mc Connell, 1967; Song & Song, 1969).

Since the current research centres on the ability of the BGT-*c* to detect brain damage and to differentiate between functional and organic impairment, literature concerning these two functions is detailed later in this section (see 2.1.9: p.24). Before this, the BGT-*c* is contextualized in its role as a screening test and the variations of the BGT-*c* are discussed.

2.1.5 The expanded BGT as a useful diagnostic screening test

It is important to emphasize that in this thesis no attempt is being made to advocate the sole use of the expanded BGT to the exclusion of other, more comprehensive measures. Instead, it is suggested in the course of this study that the expanded BGT is a valuable screening technique to be used in the initial assessment of individuals in order to identify the possible *presence* of brain damage. Lezak (1983, p.152) notes that screening "operates only as an early warning system". In contrast, neuropsychological batteries provide more detailed information which clarifies the *consequences and extent* of neurological damage (Arbit & Zager, 1978; Russell, 1976). Nor is neuropsychological measurement being advocated as the only means of assessment. A comprehensive assessment includes history taking, observation and various medical and psychological measures (Lezak, 1983, 1989b; Kampen & Grafman, 1989). As Eno and Deichmann (1980, p.43) state: "Alone, the BGVMT, or for that matter any other psychometric instrument cannot be used to diagnose the conditions of brain damage". Similarly, Lezak (1983, p.152) points out that no neuropsychological test is fail-proof and that more comprehensive evaluations should follow if necessary.

Lezak (1983, p.148) notes that "limitations in predictive accuracy do not invalidate [screening] tests". In considering the accuracy of a variety of diagnostic measures, Satz, Fennell and Reilly (1970) compared the ability of six neurodiagnostic tools to identify brain damage, five of which involved medical procedures, and one which was a neuropsychological test. The researchers reported the following diagnostic accuracies: electroencephalogram (78%), skull X-ray (61%), brain scan (57%), arteriogram (69%) and pneumoencephalogram (76%), and the Block Rotation Test (70%). Satz *et al.* (1970) state that diagnostic errors are not unique to neuropsychological tests and concluded that no unitary method is adequate. Additionally, research notes that conducting a pneumoencephalogram is a complex and hazardous procedure (Eno & Deichmann, 1980; Satz *et al.*, 1970). Overall, the mean accuracy for all neuropsychological tests is 75%, while that of the BGT-*c* is 76% (Heaton, Lyle, Baade & Johnson, 1976). When Hutt-Briskin's 12 brain damage indicators as standardized by Lacks (1984) are used to score the test, the accuracy of the BGT-*c* is as high as 84%. This diagnostic accuracy was achieved in both identifying brain damage *and* in distinguishing between organically impaired and functionally impaired psychiatric patients (Lacks, 1984).

In addition, neuropsychological tests are increasingly being used to supplement medical tests because they have certain benefits over medical procedures in that they pose neither risk nor significant discomfort to those being tested (Heaton *et al.*, 1976). Neuropsychological tests also provide information about the patient's dysfunctions and assets that are relevant to daily functioning (Heaton *et al.*, 1976). Boll (1974) adds that neuropsychological measures have a particular value in the assessment of the behavioural variations evident in psychiatric and neurologically impaired patients.

Screening is useful to the extent to which it is able to recognize an illness or the high risk of one (such as in genetic diseases), and the degree to which it can prevent the onset or improve the prognosis of such a disease due to early treatment (Marteau, 1994). However, Lezak (1983, p.151) stated that neuropsychological screening "has only limited usefulness at present". She reasoned that due to the increased sophistication of neuropsychology, in most instances of referral for the assessment of brain functioning the presence of damage is either evident or already clearly noted. She admits that screening may still be functional when a meticulous assessment of each individual is not possible. It must be noted that Lezak (1983) held this opinion 12 years ago and in an American context. However, in the current South African context, the situation is different to that described by Lezak (1983).

There are many circumstances in which a quick method of testing for neurological impairment is needed and relevant in contemporary South African conditions. For example, broad screening becomes significant when underprivileged children enter preschool programmes (McNamara, Porterfield & Miller, 1969). This is particularly applicable to South Africa at the moment, when one of the major aims of the new government is to educate all individuals up to a minimum level of standard seven. Screening tests are used in a broad spectrum of clinical situations, including the identification of diffuse neurological impairment at the soonest possible stage (Chouinard & Braun, 1993). Barrett Jnr, Wheatley and Laplant (1982) point out that evaluating large numbers of people is also necessary in outpatient referral situations, such as at general medical hospitals. Again, this is highly relevant in South Africa where primary health care has not been accessible to large portions of the population, and the new government plans to implement accessible health care for all citizens. Screening for organic damage is also a frequent procedure in admissions to mental hospitals. In addition, mental health workers are often called upon by organizations such as courts, welfare agencies, schools and families to establish whether a specific condition is emotionally or organically implicated (Golden, 1976). The large numbers of individuals to which screening tests are appropriate render the need for a screening test which is rapid both to administer and to score.

In addition, neuropsychological screening measures are valuable for an initial diagnosis. Results of screening tests assist both in decisions regarding treatment and in decisions as to whether more intrusive or more costly measures are indicated (Chouinard & Braun, 1993). In such circumstances, the time factor often plays a key role (Golden, 1976). McNamara *et al.* (1969) comment that the BGT-*c* as a screening measure gains additional significance when the time factor is considered. Golden (1976) states that while the Halstead-Reitan Battery has an accuracy rate as high as 90% in the identification of brain damage, it takes from 4 to 12 hours to administer. The administration of this battery "entails a commitment to time that few clinicians in mental health centres are able to give" (Golden, 1976, p.821). Others endorse this opinion (Barrett Jnr *et al.*, 1982; Russell, 1976). Also, reliable interpretation of this battery requires a high degree of specialization "impossible to a practitioner already involved in many other activities" (Davidson, 1974, p.11). According to Barrett Jnr *et al.* (1982), another limitation is the cost of the complex equipment needed for the administration of this battery, who question the expense of applying this battery to each individual referred for assessment. While the Halstead-Reitan Battery is indicated when diagnostic accuracy needs to be extremely precise or when the actual *extent* of damage needs to be evaluated, the BGT-*c* is beneficial as a basic screening test (Russell, 1976).

Eno & Deichmann (1980) maintain that on its own the BGT-*c* is a competent screening tool for identifying diffuse dysfunctions. Numerous researchers testify to the usefulness of the BGT-*c* as a screening measure to identify neurological impairment and to differentiate between organic and functional impairment (e.g. Arnold, Heustis, Wemmer & Smeltzer, 1978; Clark, 1982; Clarke & Leslie, 1971; Flick & Duncan, 1973; Frostig *et al.*, 1961; Koppitz, 1975; Lacks, 1984; Landis *et al.*, 1974; Racusin & Moss, 1991; Skeen *et al.*, 1982), which will be detailed later in this section (see 2.1.9: p.24). In addition to a screening measure, the BGT-*c* can also be useful as part of a battery when more specific information is required (Eno & Deichmann, 1980; Lacks, 1982).

Neurological impairment can be detected by screening tests because they either identify "a particular defect or set of associated defects, or general impairment or characteristic signs of organicity" (Lezak, 1983, p.148). Therefore, when neuropsychological tests are used as screening tools, the highest diagnostic accuracy is obtained by combining these three elements, namely by combining tests which detect a particular dysfunction, with those which detect overall dysfunctioning and others which extract "diagnostic signs" (Lezak, 1983, p.148). Specific organic indicators include perseveration which is described by Lezak (1983, p.148) as "the continuation of a response after it is no longer appropriate". Perseveration is so closely linked to neurological impairment that this trait alone should lead the practitioner to suspect impairment (Lezak, 1983). Lezak (1983) also cites research which indicates that the rotation of designs is another 'diagnostic sign'. Lezak (1983) states that immediate memory impairment is one function which is sensitive to a decline in overall intellectual processes. She adds that measures which are sensitive to diffuse intellectual impairment are accurate identifiers of neurological impairment, and produce limited false negatives. But she warns that the danger with such measures is that psychiatric patients often score false positives on these tests. It will be argued in the course of this work that the expanded BGT contains all three of these elements, namely measurements of specific functions, overall functioning and diagnostic signs.

2.1.6 Variations of the BGT-*c*

A variety of versions of the BGT-*c* have appeared over the years. As discussed (see 2.1.1: p.5), the BGT-*c* involves the copying of the BG designs as opposed to recalling them. First the Minnesota Percepto-Diagnostic Test is briefly described for historical interest only. It consists of some of the BG designs, leading this author to presume it began as a variation of the BGT-*c*. Next, certain adaptations which have been used in very few studies or in a single one only, are summarized. Subsequently, one of the better known variations, the Background Interference Procedure (BIP) which was developed by Canter (1966), is considered. Since it stimulated research interest at the time of its inception it is discussed in some detail. Finally Bender's (1938) original community of property is described.

2.1.6.1 The Minnesota Percepto-Diagnostic Test

The Minnesota Percepto-Diagnostic Test is an acknowledged neuropsychological test in its own right (Lezak, 1983). Lezak (1983) states that it can be administered to both children and adults. It utilizes two of the BG designs, specifically designs A and 4 (see Appendix 1: p.231). The designs are presented in a variety of orientations and the test is scored by analysing the rotations. Lezak (1983) describes this test as useful for differentiating between psychiatric patients with and without neurological damage as well as for identifying non-psychiatric patients with cerebral impairment.

2.1.6.2 Lesser known variations of the BGT

Lesser known variations include the "stress Bender" (Lezak, 1983, p.387), which involves a second BGT-*c*, during which subjects are asked to execute the test as rapidly as possible. While this is being done, the tester times the subjects, and draws attention to this fact by deliberately making a noise. Lezak(1983) states that individuals with mild deficits may counteract their problems when under no stress, but reveal impairment under time pressure. In contrast, numerous neurologically normal individuals actually improve their performances under these conditions (Lezak, 1983). There is little research to indicate the value of this version. In addition, it is not practical for inter-study comparisons, as the interference by the examiner would be difficult to replicate precisely and therefore it would be difficult to standardize.

A multiple-choice recognition Bender requires subjects in each question to identify which one of four designs accurately represents the designs which they copied during the BGT-*c* (Allen *et al.*, 1971). Although this method differentiated between mentally retarded and normal children, differences between the two subject groups were more significant on the BGT-*c*. In another version, Smith and Martin (1967) used various learning cues which included tracing the design from the card over a piece of copy paper. They found that the ability of subjects to correct rotations successfully differentiated between neurologically impaired and normal children.

Brannigan and Brunner (1991, p.286) developed a "modified BGT" using six of the original nine designs, which was replicated by Parsons and Weinberg (1993). Pope and Snyder (1970) introduced a variation which used only two of the original BG designs but added four modified designs. These were similar to the two original designs, but easier to draw. In another variation, Weiss (1971b) used four of the BG designs to investigate the connection between the direction in which the designs were drawn (from left to right or from right to left) and the reading habits of school children.

While the above forms of the BGT-*c* are of interest, none have fully established themselves as unequivocally useful.

2.1.6.3 The Background Interference Procedure

The Background Interference Procedure (BIP) was developed by Canter (1966) and involves a second BGT-*c*. The BIP requires that the designs are copied onto paper which is covered with thick black crossing wavy lines, introduced to add an interference dimension to the test. Performance on the first and second protocols are scored according to a modification of the Pascal-Suttell scoring system especially developed by Canter (1966) to screen for brain damage in psychiatric patients. The difference between the scores of the two tests is then computed. The idea originated from Canter's (1966) observations that patients with organic impairment experienced difficulty in copying the BG designs onto such paper. He developed this method by comparing the performance of psychotic and non-psychotic psychiatric patients without brain damage to those who did have cerebral impairment. He then identified specific errors from these protocols and used these criteria to differentiate between the presence or absence of cerebral disorders. He reported a diagnostic error of 9% and concluded that the BIP method is extremely sensitive to brain damage. However, he admitted that certain of the classification criteria need re-evaluation.

Other researchers have investigated this method with varying results. Sabatino and Ysseldyke (1972) found that the BIP successfully discriminated between children with reading deficits and those without, while the BGT-*c* did not. Yulis (1970) used the BIP on normal and brain damaged subjects but added further distractions. He concluded that the BIP, at least in part, determined motivational drive. In contrast to studies supporting the sensitivity of the BIP, other research has found it to be unsatisfactory. For example, Delaney (1982, p.834) found that the BIP could not differentiate adequately between normal and epileptic persons and described the accuracy as "disappointingly low". Boake and Adams (1982) also questioned the usefulness of the BIP after obtaining a false positive rate of 42% when using it to differentiate between cerebrally impaired and cerebrally normal psychiatric patients. They criticized the low reliability of the test and stated that the diagnostic validity of the test was lower than earlier reports suggested. They concluded that the utility of the BIP in screening for brain damage was "limiting" (Boake & Adams, 1982, p.627). Holland and Wadsworth (1979) found that the BIP method did not have a higher diagnostic accuracy than the BGRT.

The fact that some researchers used scoring systems other than that which Canter (1966) developed, confuses research further (e.g. Sabatino & Ysseldyke, 1972; Snortum, 1965; Song & Song, 1969). Song and Song (1969) used the BIP method to investigate the performances of organically retarded individuals, functionally disturbed mentally retarded psychiatric inpatients without organic impairment, and culturally-retarded individuals without evidence of impairment. They used the Pascal-Suttell scoring method. Although they found BIP performances did differentiate between organically impaired and non-organically impaired individuals, a frequency distribution exposed a substantial overlap between the three groups. This means that numerous individuals from each of the groups obtained the same scores. The BIP also failed to differentiate between the scores of the culturally and emotionally retarded groups in this study. In addition, research revealed that the BIP is prone to a range of confounding variables. For example, Adams, Boake and Crain (1982) found that the BIP has a high rate of false positives. They concluded that the test was so seriously affected by variables such as age, education, premorbid intelligence and racial-ethnic factors, that specific characteristics of subjects would result in a diagnosis of cerebral impairment, whatever the subjects' neurological condition.

To conclude, Lacks (1984) states that there is insufficient and inconsistent evidence as to the usefulness of the BIP. Research with children has also produced "inconclusive results" (Koppitz, 1975, p.117). In support of this, Delaney (1982) states that the rate of differentiation of this version is unsatisfactorily low. Further, Eno and Deichmann (1980, p.42) state that the BIP failed to increase the diagnostic accuracy of the BGT-*c* "by an appreciable amount". Considering the above information, research on the BIP appears contradictory and further investigation into this method is needed before its usefulness can be adequately determined. Further, none of the other versions discussed above represent an *improvement* on the original BGT-*c*, which Bender (1938, 1946) developed and used.

2.1.6.4 The original BGT-*c* version

Since the majority of the literature reviewed in this entire section relates to the original BGT-*c* developed by Bender (1938), it is unnecessary to discuss it in detail here. Suffice to say that the utility of the original BGT-*c* is demonstrated by extensive research over more than half a century which supports its diagnostic and differential abilities (e.g. Arnold *et al.*, 1978; Brilliant & Gynther, 1963; Clark, 1982; Clarke & Leslie, 1971; Dana *et al.*, 1983; Flick & Duncan, 1973; Frostig *et al.*, 1961; Goldberg, 1959; Griffith & Taylor, 1960; Hain, 1964; Jernigan, 1967; Koppitz, 1975; Lacks, 1984; Lacks *et al.*, 1970; Lacks & Newport, 1980; Landis *et al.*, 1974; Lezak, 1983; Lubin *et al.*, 1984; Schulberg & Tolor, 1961; Skeen *et al.*, 1982). Although the administration of the BGT-*c* varies to some extent, the basic test remains the same. It involves the copying of the nine BG designs, usually without a time limit (e.g. Koppitz, 1975; Lacks, 1984, Lezak, 1983). Details of administration will be dealt with later (see Chapter 5: p.91).

2.1.7 Functions measured by the BGT-*c*

Lezak (1983) cites research which indicates that the copying function of the BGT-*c* requires an advanced degree of integrational and organizational functioning which is more complex than motor or visual imaging processes alone. This concurs with the fundamental principles of the gestalt. As already discussed (see 2.1.2: p.6), the copied BG designs are the products of the organization which occurs in the perceptual process. Bender (1938, p.3) describes a copied gestalt design as a "visual motor pattern" which reveals changes from the original design pattern introduced by the perceptual integrating abilities of the individual executing the test. Since integrational and organizational functions are involved, the copying abilities of an individual on the BGT-*c* are predisposed to disintegrate with various types of brain impairment where these integrational and organizational functions are impaired (Bender, 1938). Koppitz (1975) states that difficulty in copying the designs is due to immaturity (and therefore to underdeveloped perceptual abilities), impaired perceptual abilities, visual perceptual impairment, poor motor coordination or a combination of these factors.

Experts vary on exactly what type of brain damage the BGT-*c* detects: McFie (1975) states that the copying ability derived from the BGT-*c* appears to involve a greater degree of left parietal constructional ability than the right parietal spatial perception, which is elicited by the Benton Visual Retention Test. The BGT-*c* is also sensitive to diffuse cortical or subcortical deterioration (Allen *et al.*, 1971). Individuals with parietal lobe lesions are likely to produce poor reproductions of the original design (Garron & Cheifetz, 1965). However, Lezak (1983) has a broader view of the BGT-*c* functions. She states that from her own experience, distortions of omission occur on the BGT protocols of individuals with right *or* left parietal damage, and especially on those with individuals who have bilateral neurological impairment. She classifies this test under the heading of "constructional functions" which are separated into "drawing" tasks and "building or assembling" tasks (Lezak, 1983, p.382). The drawing group of neuropsychological tests is further subdivided into "free drawing" and "copying" (Lezak, 1983, p.384). Therefore the BGT-*c* can be described as a constructional copy (drawing) test which is visuomotor in nature. 'Constructional' tasks involve a complex combination of functions which incorporate visual, perceptual and motor processes. In addition, constructional tasks consistently contain a spatial element (Lezak, 1983).

Lezak (1983) and Brook (1975) state that drawing tests are a favoured neuropsychological tool because of their ability to detect a broad variety of brain damage. Drawing tests are susceptible to perceptual, motor and organizational dysfunctions and to particular cognitive deficits (Lezak, 1983). However, Lezak (1983) cautions that many facets of impaired intellectual functioning are not detected by these types of tests. She notes that organically impaired individuals have been known to perform well in drawing tests although this should have been precluded by the nature of their cerebral damage. She adds that individuals who are handicapped on drawing tasks may be capable of copying accurately. She cites Messerli, Seron and Tissot (1979, in Lezak 1983) who maintain that the reverse is seldom true.

The gestalt function of perception can be further understood by considering the manner in which different but concurrent information processes are executed by the left and right brain hemispheres. Lezak (1983) describes the different functions performed by the left and right hemisphere in constructional tasks. The *left* hemisphere is responsible for the processing of "linear" information such as verbal, numerical and symbolic data (Lezak, 1983, p.57). To process information, the left hemisphere reduces visual data to components which can be verbally defined, such as the size and circumference of a circle. The *right* hemisphere function can be described as "configurational" and it mediates data which cannot be accurately defined verbally or symbolically, such as facial recognition and three dimensional percepts (Lezak, 1983, p.57). Copying and drawing functions are related to the right hemisphere. In contrast, the right hemisphere processes visual information as "spatially related wholes" and, as such, is responsible for pattern recognition and recall (Lezak, 1983, p.56). Lezak (1983) states that, despite the above explanation being too simplistic, it nevertheless illustrates how the two hemispheres function together in the processing of information.

Considering that the proportions and spacial aspects are important in the reproduction of the BG designs, left hemisphere functioning is clearly necessary for accurate copying. Further, the role of the right hemisphere regarding pattern recognition and organization is also important in the light of gestalten being perceived as patterns, as already discussed (see 2.1.2: p.6). From this explanation the degree of integration needed to process gestalt configurations becomes clearer, especially since it appears that both cerebral hemispheres are involved.

Lezak (1983) explains that in constructional tasks, individuals with left hemisphere damage may distort designs because they are inclined to simplify these. They may also find drawing angles to be problematic. Individuals with right hemisphere damage experience problems with copying designs because their perceptual organizational abilities are likely to be impaired, which creates difficulty in integrating complex information. This concurs with the opinion of this author expressed above, that the processing of gestalt configurations involves both hemispheres. It also supports Bender's (1938) early predictions about the usefulness of the test as well as Lezak's (1983) view that damage to the left and right hemispheres can be detected by the BGT.

2.1.8 The importance of the identification of brain damage and the need for differentiation between functional and organic impairment

2.1.8.1 The need to identify neurological damage

Accurate, rapid and early assessment to identify the *presence* of brain damage is important for a number of reasons. These include the prevention of further, secondary damage (Levin, Eisenberg & Miner, 1983; Pang, 1989), facilitation of treatment (Gronwall, 1989; Lezak, 1983) and rehabilitation (Gronwall, 1989; Lezak, 1983). Assessment also indicates a prognosis and alerts medical specialists to the extent of damage, and to problem areas (Gronwall, 1989; Lezak, 1989a). It is particularly vital for those whose injuries are considered mild as it obviates the many psychosocial consequences in the family and occupational life of an individual in whom these deficits are unidentified and therefore unsupported (Gronwall, 1989). Further, Vakil, Arbell, Gozlan, Hoofien & Blachstein (1992) stress the pivotal role which memory dysfunction has on rehabilitation. Also, rapid assessment allows the "patient's potential to rehabilitation [to be] maximised" (Lezak, 1989a, p.xiii).

In addition, providing information from an assessment to the victim and close relatives allows those involved to make the necessary adaptations to their practical and psychological framework to best cope with the situation (Gronwall, 1989). While cognitive impairment affects the quality of life of the sufferer and the family, counselling and an understanding of the impairment is beneficial and reduces distress (Gronwall, 1989; McIntosh-Michaelis *et al.* 1991). Meltzer (1983), a well known psychologist who suffered from memory impairment following brain damage caused by anoxia after a heart attack, stresses the need for precise information. For the above reasons, accurate, rapid assessment is of vital importance.

Problems in the assessment of brain damage include the detection of relatively mild injuries, the assessment of psychiatric patients and the detection of malingering (Gronwall, 1989; Lezak, 1983). In addition, neurologically impaired individuals often have concentration problems (Lezak, 1989a) and tire easily (Gronwall, 1989; Haaland, Temkin, Randahl & Dikmen, 1994). Clearly a long battery of tests is contra-indicated with such subjects.

2.1.8.2 The need for differentiation between functional and organic impairment

In clinical assessment, the identification of brain damage alone is insufficient for diagnostic and treatment purposes. As already noted (see 2.1.8.1: p.22), assessing brain damage in psychiatric patients can be problematic. Lezak (1983) states that for neuropsychological tests to be clinically valuable they must be able to distinguish between individuals whose presenting symptoms are difficult to diagnose in other ways. Such differentiation is important, because the presenting symptoms of organic and functional impairment can be very similar, to the point of mimicking one another. For example, a psychotic disorder due to a general medical disorder or a substance-induced psychosis may imitate schizophrenia, and amnesia due to a general medical condition is symptomatically comparable to dissociative amnesia (American Psychological Association, 1994; Coleman *et al.*, 1984;

Kaplan & Sadock, 1991). Cronbach (1991) adds that behavioural disorders can have an organic rather than an emotional base and Levin *et al.* (1983) note that behavioural problems occur frequently in head-injured children. Therefore, *differentiation* is of the utmost importance for various clinical decisions, the most important of which involves the course of treatment chosen. On this topic, Strain and Kinzie (1969) note that clinicians often experience problems in diagnosing brain damage from neuropsychological tests when the sample includes schizophrenics. They report that with the Halstead-Reitan Battery, a high number of scores of schizophrenics fall within the range which is indicative of neurological impairment. This emphasizes the need to reduce the misdiagnosis of psychotic individuals on tests for brain damage (Strain & Kinzie, 1969).

2.1.9 The ability of the BGT-*c* to detect neurological damage and to differentiate between functional and neurological impairment

As already mentioned (see Chapter 1: p.1), a variety of scoring systems exist for the BGT-*c*. The diagnostic accuracy of these scoring systems appears to vary. Lubin and Sands (1992) note that the literature concerning the BGT-*c* should be considered in the light of these variations. For this reason, where appropriate in this sub-section, specific references are made to the scoring systems which have been successfully used.

2.1.9.1 The BGT-*c*'s ability to identify brain damage

The BGT-*c*'s ability to detect brain damage is well established by research (see 2.1.4: p.10). This review describes a single, rigorous study in detail, followed by brief references to other studies.

An important study regarding the construct validity of the BGT-*c* was conducted on open heart surgery patients. Landis *et al.* (1974) investigated the extent of organic damage incurred by cardiopulmonary bypass apparatus used during the operation. The researchers state that, in the process of redirecting the blood to bypass the heart and lungs, the apparatus creates tiny particles (microemboli), which are implicated in both central nervous system and neurological impairment. In addition, this study used the BGT-*c* to establish the efficacy of a new filter designed to filtrate these microemboli. Subjects were administered the BGT-*c* pre- and post-operatively. The Pascal-Suttell scoring system was used to evaluate the protocols. The researchers found that subjects on whom the filter had been used revealed no deterioration on the BGT-*c*, while the protocols of patients on whom no filters had been used recorded a 57% deterioration. Interim scores indicated that various levels of deterioration were concomitant with the application of varying degrees of microembolic filtering. This study also reaffirmed the validity of the BGT-*c* since the various BGT-*c* scores concurred with microembolic counts made using ultrasonic equipment.

The ability of the BGT-*c* to detect neurological dysfunction is supported by other research (e.g. Bender, 1946; Burgess, Kodanaz, Ziegler & Greenburg, 1970; Lezak, 1983; Russell, 1976; Yulis, 1970). Research investigating the capacity of the BGT-*c* to measure the *extent* of damage has also been done. Cooper, Dwarshuis and Blechman (1967) tested 40 subjects with brain damage who were classified into three categories - namely mildly impaired, moderately impaired and severely impaired. The test results for each group correlated significantly with the degree of impairment. Since this test is able to establish degrees of neurological impairment, it has been used to gauge improvement or deterioration in neurological functioning (e.g. Bender, 1946; Landis *et al.*, 1974). It is used to monitor deterioration from diseases which implicate the central nervous system, for example, the sickle cell trait in children (Flick & Duncan, 1973).

The BGT-*c* has also been useful in the identification of brain damage in children. An early study by Koppitz (1962) used a large sample of 384 children, of which about 33% had been diagnosed as neurologically impaired. A diagnostic accuracy of over 91% was reported. Smith and Martin's study (1967) supported these findings. They used a variation of the BGT-*c* in a study of 50 children and reported a diagnostic accuracy of approximately 90%. Other studies have corroborated the above research findings (Holroyd, 1966a; Wagner & Murray, 1969). Studies conducted on retarded children have further supported the above research. Condell (1963) found that the test successfully discriminated between three levels of retarded children as well as between retarded and normal children. Allen *et al.* (1971) found a highly significant difference in the abilities of educable mentally retarded and normal individuals to reproduce and recognize the BG designs.

2.1.9.2 The BGT-*c*'s ability to differentiate between functional and organic impairment

The BGT-*c* can differentiate between similar behavioural conditions by distinguishing between organic and functional impairment. Many studies on this topic exist. For the purposes of this research, major studies of definitive research value are discussed in some detail, while lesser or similar studies are mentioned briefly.

Some of the most definitive and rigorous studies documented on the BGT-*c* have been conducted by Lacks on her own or in conjunction with other researchers (Brilliant[1] and Gynther, 1963; Lacks *et al.*, 1970; Lacks & Newport, 1980; Lacks, 1984). For example, Lacks and Newport (1980) reported an 84% diagnostic accuracy in distinguishing between organically impaired and functionally impaired psychiatric patients. Hutt-Briskin's 12 brain damage indicators which were operationalized by Lacks (1984), were used to evaluate the protocols. This rate of diagnostic accuracy is higher than the average accuracy for neuropsychological tests of 75% and is also higher than the average diagnostic rate of the BGT itself, which is 76% (Heaton *et al.*, 1976). Lacks and Newport (1980) also reported a high inter-scorer reliability. The organic presence in the subjects was approximately 34%, which, according to Heaton *et al.* (1976), is representative of most populations of psychiatric neuropsychological referees.

Earlier studies conducted by Lacks have had equally successful results. A study by Brilliant and Gynther (1963) examined three tests used to discriminate between organic and functional impairment namely the BGT-*c*, The Benton Visual Retention and the Graham-Kendall Memory-for-Designs. The sample group of 120 individuals included those with diagnoses of both chronic and acute brain syndrome, alcoholism, personality disorders and psychoses. The BGT-*c*, scored according to the Hutt-Briskin method, correctly diagnosed 82% of the sample. The researchers concluded that this test was the "best single measure" for identifying brain damage in a psychiatric sample (Brilliant & Gynther, 1963, p.474). In another study which uses Lacks's adaptation of the Hutt-Briskin indicators, Lacks *et al.* (1970) compared the efficacy of the Halstead-Reitan Battery with that of the BGT-*c*, using

---

1. Lacks's maiden name was Brilliant.

a sample of 64 hospitalized organic, schizophrenic and medical patients. The researchers reported that the Halstead-Reitan Battery takes up to four hours to administer, while the BGT-*c* takes approximately five minutes (Lacks *et al.*, 1970). Both procedures correctly *identified* patients with neurological impairment, with a high degree of accuracy. The researchers concluded that the use of the Halstead-Reitan Battery to screen for neurological impairment can "be seriously questioned" as it was no more effective than the brief BGT-*c* (Lacks *et al.*, 1970, p.481). Lacks *et al.*'s (1970) study (1963) further endorsed the value of the BGT-*c* as it *differentiated* between non-organic, normal and schizophrenic patients with a 91% accuracy. The lengthy Halstead-Reitan Battery's differentiating accuracy was only 62%. When using Lacks's scoring system, Marsico and Wagner (1990) found a diagnostic accuracy among psychiatric outpatients of 78,8% and 81,3% when using the Pascal-Suttell scoring system.

Many other studies have investigated the differential diagnostic abilities of the BGT-*c* on samples of psychiatric patients. For example, Holland and Wadsworth (1979) found that the test significantly differentiated between organically impaired and schizophrenic inpatients, regardless of age, IQ or socioeconomic status. Further evidence of the ability of the BGT-*c* to discriminate between psychiatric patients was established by Song and Song (1969). They tested 87 patients who were categorized as functionally impaired, organically impaired, and non-organically retarded (due to non-organic causes such as cultural and family conditions). Although the test did not differentiate between the cultural-familial disadvantaged and the functionally impaired groups, it separated the organic and non-organic groups at the 5% level of significance.

Similar findings among both psychiatric inpatients and outpatients have been found by other researchers. Marsico and Wagner (1990) reported a diagnostic accuracy of 74% on a sample of 104 individuals using the BGT-*c* and 73% when using a single BG design. Rosecrans and Schaffer (1969) tested a sample of 77 individuals which consisted of two groups suffering from neurological impairment, but which differed in age. The younger group had a mean age of 36,5 years and the older group an age mean of 63,1 years. The third group consisted of non-organic, mixed psychiatric patients. The researchers concluded that the BGT-*c* scores were "highly efficient discriminators" between the organic and functionally disordered groups (Rosecrans & Schaffer, 1969, p.410). Griffith and Taylor (1960) found that a large sample of 1000 protocols significantly indicated the difference between functionally impaired subjects and organically impaired patients. Goldberg (1959) reported an 82% diagnostic accuracy using the BGT-*c* to differentiate between similar groups. The BGT-*c* was also able to identify neurological impairments due to heroin abuse in both non-psychotic addicts and psychotic addicts (Korin, 1974).

In addition to the above studies which compare only impaired individuals, a number of studies have included normal subjects in their samples. For example, Hain's (1964) study is one of the earliest, rigorously conducted studies to investigate the differentiating efficacy of the BGT-*c*. His sample of 93 individuals included normal, neurologically impaired and psychiatric individuals. Using a scoring system he developed, Hain (1964) found a highly significant difference between individuals with and those without brain damage. Further, supporting the discriminatory ability of the BGT-*c*, Bowland and Deabler (1956) reported that both qualitative and quantitative evaluations of BGT-*c* protocols were able to differentiate between the three groups, namely normal, brain damaged and psychotic individuals, at a level above chance expectation. Korman and Blumberg (1963) differentiated between the three groups using the BGT-*c*, scored according to the Pascal-Suttell method, with an overall diagnostic accuracy of 73%. Levine and Feirstein (1972) used the Hutt-Briskin signs converted into standard scores, and found significant differences between organic and non-organic groups, but not between the normal and schizophrenic groups.

The BGT-*c* has also successfully differentiated between emotional and organic pathologies in children, although the published literature concerning younger individuals is considerably less than that available on adult studies. Culbertson and Gunn (1966) appear to have conducted one of the earliest of such studies. Their sample of 65 children included individuals who were emotionally disturbed, schizophrenic, retarded and those suffering from organic brain syndrome. Although not statistically rigorous because they used the Koppitz Developmental Scoring System beyond the specified age range (see 3.3.4: p.67), they reported the lowest scores in the organic group. Mc Connell (1967) studied a mixed group of 120 children with either emotional or organic impairment. This ambitious study aimed at not only differentiating between organic and functional disorders, but also between the various *levels* of involvement of both impairments. The three levels of organic involvement were considerable, minor and none. Emotionally dysfunctional individuals were classified into four levels - namely "situational", "neurotic or characterlogical", "borderline psychotic" and "psychotic" individuals (Mc Connell, 1967, p.371). The researchers used three different scoring systems to evaluate the protocols. The test documented highly significant differences between all groups of neurologically impaired individuals from all emotionally disturbed groups, with one exception. It was found that non-organic psychotics did not differ significantly from the group with minimal brain impairment. Maloney and Ward (1970) also investigated the degree of impairment in mild to severely retarded adolescent inpatients. They reported that the BGT-*c* differentiated between organically and functionally retarded individuals with a high inter-scorer reliability.

With regard to the difference between normal and brain impaired or retarded children and the severity of their impairment, Wagner and Murray (1969) concluded from their research findings that the BGT-*c* appears to be a cogent indicator of organic damage in children and also a differentiating device. Other studies confirm this research (Allen *et al.*, 1971; Flick & Duncan, 1973; Hutton, 1966; Maloney & Ward, 1970; Mc Connell, 1967). This differential ability is of importance as 75% of retardation is regarded as having a sociocultural aetiology (Coleman *et al.*, 1984).

In addition to the usefulness of the BGT-*c*, as discussed above, this test has a number of advantages. However, certain problems also emerged, following a detailed consideration of the literature. The next section critically evaluates the BGT-*c*.

2.2 A critical evaluation of the BGT

In this evaluation, supplementary assets of the BGT-*c* are discussed and include the potential introduced by group testing. Next, certain problems evident in the BGT-*c* are elucidated upon, including the effect of education on this test and the absence of normative data for low-educated populations. Criticism of the BGT-*c* is also noted. The section concludes with the argument that, despite its shortcomings, the BGT-*c* is a valuable screening measure.

2.2.1 Supplementary attributes of the BGT-*c*

Apart from its established use as an indicator of brain damage (see 2.1.9.1: p.24) and its ability to differentiate between functional and neurological impairment (see 2.1.9.2: p.26), the BGT-*c* has a number of inherent supplementary attributes which enhance its worth.

2.2.1.1 The broad applicability of the BGT-*c*

The BGT-*c* can be used on both adults and children, making it a test which is economical to buy and to learn to administer and score. Since the test involves copying designs, no writing skills are necessary, which makes it useful for assessing individuals with very little or no education, whether due to lack of schooling, retardation or because they are very young. For this reason, the BGT-*c* is frequently used in the initial assessment of patients admitted to mental hospitals who have little or no education. It has a low verbal content and the administration instructions are simple, which is a further advantage in a country such as South Africa in which a large number of different languages exist.

2.2.1.2 Rapid administration time appropriate for difficult testing situations

The rapid administration time of the BGT-*c* facilitates its use in circumstances where longer tests would be inappropriate. In this regard, Landis *et al.* (1974) found it viable to administer to postoperative open heart patients. The physical and mental condition of such patients and their location in a surgical ward was "not conducive to sustained psychological evaluation" and concentration for more that 20 minutes would have been extremely trying (Landis *et al.* 1974, p.558). Also, Lezak (1983) notes that individuals with neurological impairment quickly become fatigued. Thus, the short administration time of his test renders it appropriate for use with subjects who tire easily or who have a limited attention span, such as very old or ill people as well as hyperactive psychiatric patients and individuals with cerebral damage.

2.2.1.3 Group versus individual testing

The group testing method of the BGT-*c* is regarded by Koppitz (1975) as the single most significant recent breakthrough in research on the BGT. She stated that it was first published by Keogh and Smith (Koppitz, 1975). Yet a study conducted as early as 1956 by Bowland and Deabler also used this method. However, the Keogh and Smith (1961) study appears to be the first research conducted to determine whether the results of the group administration differed significantly from those of the individual method of administration. In this study, two different scoring methods and two modes of group administration were used. No significant differences between the group and individual methods were found. Later studies endorse the group administration method as equally reliable and valid as the individual method (Becker & Sabatino, 1971; McCarthy, 1975). The group method has been used successfully with both children and adults, as well as with mixed sample groups of normal, organically impaired and psychiatric patients. (Becker & Sabatino, 1971, Bowland & Deabler, 1956; Dibner & Korn, 1969; Keogh, 1965; Keogh & Smith, 1961; McCarthy, 1975; Viljoen, Levett, Tredoux & Anderson, 1994). Group size varies from 2 to 30 (Koppitz, 1975, McCarthy, 1975) and some groups have included entire school classes with no actual numbers given (Viljoen *et al.*, 1994).

The group testing method is highly appropriate to the time constraints experienced by clinicians, particularly in the South African context, where the ratio of clinical psychologists to the overall population is as low as 1:35 800 (Kriegler, 1993). This method increases the number of subjects which it is feasible to test, resulting in larger sample numbers which otherwise would be difficult to achieve. Increased numbers increase the reliability and validity of the normative data obtained. For example, Becker and Sabatino (1971) calculated that the administration and scoring of the group test took a third of the time that the individual method would have taken. They tested a sample of 169 children. Since the group administration method does not compromise the test's validity, it is a viable alternative to the individual testing method.

2.2.1.4 High test-retest reliability and detection of malingering

Another important aspect of the BGT-*c* is its high test-retest reliability. Goff and Parker (1969) report a significant test-retest reliability of the BGT-*c* in children on a re-administration of the test two weeks after the initial administration. Similarly, Bowland and Deabler (1956) found no significant differences between performances of normal, psychotic, nonpsychotic and organically impaired individuals, on two administrations of the BGT-*c* spaced 2,5 hours apart. Miller and Hutt (1975) investigated the BGT-*c*'s reliability on a second administration, with schizophrenic patients, after a two-week interval. They reported a test-retest reliability of *rho* = 0,87 for males subjects and *rho* = 0,83 for female subjects, with both coefficients significant at a 1% level.

This high test-retest reliability augments the application of repeated evaluations, which is specifically useful in longitudinal studies. Longitudinal studies can be implemented to monitor diseases which implicate neurological deterioration. For example, Ris and Noll (1994) emphasize the need for neuropsychological tools which are applicable to longitudinal research, in order to establish the outcome of radiation therapy on paediatric brain tumour patients. Longitudinal studies are necessary because the neurological reaction to intercranial tumours is a "dynamic process" (Ris & Noll, 1994, p.37). Insight into these reactions can only be gained by meticulously designed prospective studies which monitor the changes in the processes as they occur (Ris & Noll, 1994). Another valid application of the BGT-*c* test-retest ability is the monitoring of improvement of neurological functioning in individuals, such as that done by Landis *et al.* (1974) in post-operative heart surgery patients.

The BGT-*c*'s value is further enhanced by its robust ability to detect malingering. Lezak (1983, 1989a) notes that malingering is problematic in neuropsychological assessment since many conditions of the brain are subtle. This is compounded by the financial compensation which is often available to those who suffer impairment (Lezak, 1989a). Research found that the BGT-*c* discriminated between organic impairment and malingering with an accuracy of 89% (Bruhn & Reed, 1975).

2.2.1.5 Economical to use

Perhaps the greatest supplementary characteristic of the BGT-*c* is its low overall cost. Two types of costs are under consideration here. The first is the total financial cost incurred in obtaining the test results. The second aspect involves the risk factor and the cost of errors.

Regarding the aspect of financial cost, Dana *et al.* (1983, p.81) aptly state that "The strength of the instrument [the BGT-*c*] lies in the cost/benefit ratio. It is brief, economical, and has potential richness of data for interpretation". The materials needed for testing are inexpensive and the rapid administration and advantage of group testing keeps expenses low. This makes the BGT-*c* a cost effective neuropsychological assessment tool and more economical than traditional medical manners of detecting neurological impairment, such as the CAT Scan or MRI. The economy of the BGT-*c* is particularly applicable in a context such as South Africa, where limited mental health funding exists.

The second cost factor, that of risk and cost of errors, must also be considered. Administration of the test poses no risk. The BGT-*c* is neither physically intrusive nor threatening, which promotes the use of neuropsychological tests as already discussed (see 2.1.5: p.12). In reference to the cost of errors, Rimm (1963) offers a comprehensive discussion of the cost of assessment procedures. He pertinently points out that the cost of a false positive or negative depends on the application of the information. The incalculable cost of not detecting cerebral damage and therefore allowing it to remain untreated or inappropriately treated must be considered. The stress of behavioural changes which are not understood has a high emotional cost, both for the families and the victim (see 2.1.8.1: p.22). The second cost of errors is the cost of false positives, or identifying brain damage, when in fact it does not exist. Rim (1963, p.91) concludes that "the cost of administering a screening test is negligible". Clearly a test with a high diagnostic accuracy reduces the number and therefore the cost of errors.

2.2.2 Problems with and criticisms of the BGT

Despite its popularity and usefulness, the BGT-*c* is controversial and has faced much criticism.

2.2.2.1 Lack of standardization

The standardization of the BGT-*c* is impaired by many deviations from the original test. These diversions include changes in the number of designs (Marsh, 1972; Parsons & Weinberg, 1993; Pascal & Suttell, 1951; Wagner & Marsico, 1991) and the design composition (e.g. Bender, 1946; Field *et al.*, 1982; Pascal & Suttell, 1951). The diversification continues into the methods of administration (Lezak, 1983; Smith & Martin, 1967; Yulis, 1970) and the versions of the tests (see 2.1.6: p.16). Not even the nomenclature of the test has avoided inconsistencies, as already discussed (see 2.1.3: p.8). The status of the BGT-*c* is further confused by a prolific conglomeration of scoring systems (e.g. Koppitz, 1975; Pascal & Suttell, 1951; Pauker, 1976). This makes the direct comparison of studies using different scoring systems difficult. It also renders normative data invalid beyond the particular scoring system for which they were established. The result of these inconsistencies is that there are many incongruencies and antitheses in both research and opinions surrounding the value of the test (Dana *et al.*, 1983). Dana *et al.* (1983, p.81) conclude that "there is no single test entity properly called the BGVMT since the designs themselves vary, standard and consensual procedure is lacking, and many different forms are used". The above deviations inhibit standardization and therefore inter-study validity and inter-study comparisons.

2.2.2.2 Education

Another problem with the BGT-*c* is the lack of normative data for low-educated populations. Yet, research unequivocally confirms that educational levels significantly influence the scores of neuropsychological measures (Adams *et al.*, 1982; Bornstein, 1986; Cronbach, 1991; Ernst, 1987; Finlayson *et al.*, 1977, Kennedy, 1981; Leckliter & Matarazzo, 1989; Prigatano & Parsons, 1976). Research has established that education is positively related to performance on neuropsychological tests, even where functions are apparently unrelated to academic performance, such as in immediate memory tests (Lezak, 1983). In a study of illiterate and low-educated older adults (a topic which appears scarce in research literature), Grossi *et al.* (1993) found a convincing educational effect on neuropsychological tests which revealed that individuals with no more than 3 years of schooling performed significantly better than those without any schooling at all.

On the BGT-*c*, research concurs with the above findings and reveals that scores are influenced by educational levels, in both children and adults (e.g. Adams *et al.*, 1982; Keogh & Smith, 1968; Koppitz, 1975). Keogh and Smith (1968) tested preschool children and stressed that the level of education needs to be considered in the interpretation of the BGT-*c*. The study by Viljoen *et al.* (1994) conducted on Zulu-speaking South African children did not specifically examine education effect. However, the reason for this was that age and education were highly correlated *r*=0,873; p=0,001) and age accounted for 31% of score variance. During a child's years of education, age and education are obviously highly correlated. Frey and Pinelli Jnr (1991), examined age and education effects and found that school grades had a more significant effect on BGT-*c* scores than age. Education effects in adults have also been found (Carlson, 1966; Lacks, 1984; Smith & Winnick, 1979). Pascal and Suttell (1951) found that their scoring system produced no education effect among non-patient adults for protocols within their stipulated age range of 15-59 years old *provided* that subjects have a minimum of 9 years of schooling. Lacks (1984) found a significant education effect, confirming an earlier study (Brilliant & Gynther, 1963), and she stratified her normative data for education. Lacks's (1984) educational levels range from a category of 3 to less than 9 years of education to a category of more than 15 years of education. The overall mean score of Lacks's (1984) lowest educational category is 3,20 errors, while that for the most highly educated category is 1,02 errors. The correlation between education and performance on the BGT-*c* reported by Lacks (1984) was r=0,51 (p<0,01).

Lacks (1984) also reports an inverse relationship between BGT-*c* performance and education and IQ variables. Error scores increase as educational and IQ levels decrease. These effects are highly significant for the sample of subjects being tested in this study as Lacks (1984, p.63) states:

> However, this effect [education] appears to be only clinically significant with those with very low education (i.e., less than eighth grade) or IQ. Even then, only 16 of the 75 or 21% of individuals with eight or fewer years of education had scores in the organic diagnostic range. It is entirely possible that at older ages or with low education or IQ there are some persons who are more likely to have some type of previously undetected brain dysfunction.

Yet despite the fact that education influences scores on the BGT-*c* and, as already discussed (see 2.2.1.1: p.30), the test is highly appropriate for testing low-educated subjects, no normative data are available for such a population group. In fact, no studies of the BGT-*c* concerning low-educated adults could be found by the present researcher, apart from Lacks (1984). Since educational effects appear substantial, it seems both appropriate and urgent that scores for low-educated individuals be investigated. This need is enhanced by the fact that the BGT-*c* lends itself to the evaluation of illiterate and low-educated adults and is consequently often used with such individuals.

2.2.2.3 Criticism of the BGT-*c*

Despite the strength of research supporting the BGT-*c* as an indicator of brain damage and a differentiator between neurological and functional impairment (see 2.1.9: p.24), it has also been severely criticized for its false positives and false negatives and for only detecting a limited number of functions (Adams *et al.*, 1982; Bigler & Ehrfurth, 1981; Butler *et al.*, 1976; Hain, 1964; Mosher & Smith, 1965; Pascal & Suttell, 1951). Margolis, Williger, Greenlief, Dunn and Gfeller (1989) found that false negatives increased as the degree of impairment worsened. Butler *et al.* (1976, p.280) noted that the test failed to detect "borderline" cases of cerebral impairment and particularly epilepsy. Shaw and Cruickshank (1956) and Delaney (1982) also found a deficit in the ability of the BGT-*c* to diagnose epileptics. Cronbach (1991) and Lezak (1983) caution against sole dependence on this test as such errors are expensive and other methods such as brain scans and syndrome-related multi-test psychometric modes of assessment are significantly more reliable. However, as already stated (see 2.1.5: p.12) sole use of the BGT-*c* is not being advocated in this thesis. Butler *et al.* (1976) support this warning and maintain that this test only detects a single facet of brain damage and its validity is poor. Garron and Cheifetz (1965) state that the BGT-*c* does not necessarily detect left frontal lesions and warns that an acceptable BGT-*c* protocol does not entirely eliminate the possibility of brain damage. Weinstein and Johnson (1964, p.820) criticized the BGT-*c*, from results of a differential study, as being of "only dubious value" for discriminating between functional and organic impairments. Weinstein and Johnson (1964, p.820) concluded that the BGT-*c* as a measure to differentiate between functionally and neurological impaired individuals, would "best serve the clinician by its deletion from his diagnostic armamentarium". However, these researchers used two methods to interpret the protocols which will be shown to be unsatisfactory in the identification of brain damage (see Chapter 3: p.57).

2.2.3 Summary

To summarize literature reviewed on the BGT-*c*, research shows it has been useful in a variety of studies and particularly in indicating neurological dysfunction and for differentiating between this and organic impairment. The diagnostic accuracy is reported to be as high as 84% when Lacks's scoring system is used. As a visual perceptual motor screening test, it measures constructional abilities and integrational perceptual organization. It appears to be able to detect damage in both left and right cerebral hemispheres. Supplementary attributes include the broad age and education range to which this test can be applied. The rapid administration makes it appropriate to those individuals who find long periods of concentration difficult. Also, the broad applicability and quick administration render it economical to busy clinicians. The group testing option further enhances its economic viability. The BGT-*c* has a high test-retest reliability and is able to detect malingering. It is also safe to use.

On the other hand, criticisms of the test include the lack of standardization in many facets including scoring and administration. In addition, some versions of the test produce a high rate of false positives and/or false negatives. Although it is highly appropriate to administer to low-educated individuals, no normative data for such populations exist. Yet literature reviewed here indicates that education exerts a significant effect on the scores obtained. The BGT-*c* measures only specific functions and other aspects of neurological functioning remain untested. The criticism that the BGT-*c* measures only one facet of brain damage is valid and this needs to be addressed. Since brain damage has a multifaceted presentation, often first manifesting with difficulties other than those tapped by the BGT-*c*, such as memory impairment, the expansion of the BGT-*c* into a memory test could significantly enhance its usefulness. The relevance and rationale of the BGRT is first investigated by exploring the role of memory in the identification of neurological impairment in the next section.

2.3 The role of memory in the assessment of brain damage

In this section, literature examining the role of memory in the assessment of brain damage is presented. The models of memory are briefly discussed followed by literature which supports the diffusion of memory functions. Next, the processes involved in memory are presented. Then studies are reviewed which firmly stress the importance of memory as a primary indication of neurological impairment, regardless of the cause. The types and duration of memory are briefly considered followed by a discussion of issues pertinent to the measurement of memory. This section ends with the argument that the expansion of the BGT to include an BGRT-*i* and BGRT-*d* will render the test more sensitive as a screening measure for cerebral impairment.

2.3.1 Theoretical models of memory

Lezak (1983, p.23) refers to "the fluidity of theoretical models of perception and memory". She quotes a paper presented by Shiffrin (1973) regarding memory theories which states that there appears to be "exactly as many models of short-term memory as there are researchers who have published their theoretical views" (Lezak, 1983, p.23). This is endorsed by the diversity of available literature concerning memory. The complexity of the subject is increased because many memory models often share analogous concepts but use "different and diverse" terminology (Reeves & Wedding, 1994, p.2). This diversity and complexity of memory models is not surprising. As already discussed (see 2.1.2: p.6), gestalt psychology adherents maintained that inferences based on the observation of perceptual experiences were necessary since the functioning of the central nervous system cannot be directly observed. (Köhler, 1920, 1947; Petermann, 1932; Wertheimer, 1922, 1923, 1925). Although knowledge of neurological functioning has increased exponentially since this time, the fact remains that intellectual functioning is not directly observable and must be understood through inferences of behavioural observation. Lezak (1983) and Kolb and Whishaw (1990) support this view. Gronwall (1989) states that the extent of damage in closed head injury can only be ascertained by the consequences. Loewenstein, Argüelles, Argüelles & Linn-Fuentes (1994, p.263) add that neuropsychology remains vital because the consequences of tumours can only be "quantified" using neuropsychological measurements. Memory is divided into numerous categories and subcategories, according to duration and type. However, the central issue in this research is the relationship between neurological damage and memory impairment.

2.3.2 Memory as a diffuse type of functioning

The debate on whether a single memory system or more than one system existed began in the late 1950s and early 1960s as researchers of both long and short term memory established contact (Baddeley, 1987, p.33). The concept that memory could involve more than an isolated unitary system was first proposed by Lashley in 1950 (Kolb & Whishaw, 1990; Pribram, 1986; Rahmann & Rahmann, 1991). The diffuse nature of memory as a combination of intricate functions and not a single function is well supported (e.g. Baddeley, 1987; Kolb & Whishaw, 1990; Lezak, 1983; Reeves and Wedding, 1994). In this regard, Rahmann and Rahmann (1991, p.266) suggest that "storage" of specific memories takes place in particular related areas of the cortex along with subcortical areas such as the hippocampus and the amygdala. They conclude that since these areas are extensive, the storage of each memory is distributed over a broad region of "neuronal networks" through the central nervous system in a manner which is "layered" (Rahmann and Rahmann, 1991, p.267). While the neuroanatomical process of memory continues to be debated, the importance of memory cannot be ignored and more is understood about the processes according to which memory occurs.

2.3.3 The process of memory

Lezak (1983, p.18) states that behaviour comprises three processes namely "intellect" (the management of information); "emotionality" (emotions and drives) and "control" (outward manifestations of behaviour). Although neurological damage usually affects all three systems, intellectual processes are highlighted in neuropsychology because: (i) the intellectual functioning of neurologically impaired individuals can be very pronounced and (ii) it is accessible to measurement (Lezak,1983). Memory and learning are one of the four major classifications of intellectual functioning and concern the manner in which information is "stored and recalled" (Lezak, 1983, p.20). Each of these four categories of intellectual functioning contain a theoretical subdivision of functions which are verbally or representationally based and those which involve nonverbal (e.g. visual or auditory) information.

Lezak (1983) states that memory and learning is fundamental to all intellectual activities. Information processing begins with sensory stimulation, followed by perception, which she describes as "the integration of sensory impressions into psychologically meaningful data" (Lezak, 1983, p.24). The perceived information is then converted into memory (Lezak, 1983, p.24). The broad cortical diffusion and intricacy of perceptual functions operative in the memory process, makes these facets susceptible to brain damage (Lezak, 1983). Organically-based perceptual impairment can occur either indirectly, through *impaired* *sensory* input, or directly, through damage to particular *integrative* formations. Perceptual integration activities consist of, *inter alia*, "awareness, recognitions, discriminations, patterning and orientation" (Lezak, 1983, p.24). Thus perception and memory are closely interrelated. Memory has four fundamental processes; registration (encoding), storage, consolidation and retrieval, and memory deficits can be caused by impairment in one or more of these functions (Lezak, 1983, Rahmann & Rahmann, 1991; Reeves & Wedding, 1994, Williams, 1979). Again, neuropsychological theories vary greatly as to which process is impaired in the case of specific memory deficits, but this is beyond the scope of this thesis.

2.3.4 Impaired memory as indicative of brain damage

Diffuse brain damage has a multitude of causes both external and internal in origin, which include infection, anoxia, hypertension, intoxications, certain degenerative, metabolic and nutritional diseases and head injuries, both open or closed (APA, 1994). Brain damage can be acute and rapid in nature (as in meningitis) or slow and chronically degenerative (as in Alzheimer's disease). Yet, regardless of the cause, the behavioural manifestations of neurological impairment usually include memory (Ellis & Young, 1988; Kolb & Whishaw, 1990; Lezak, 1983; Lishman, 1978; McFie, 1975; Oddy, 1984; Walsh, 1987; Walsh, 1991; Williams, 1979; Williams, 1991). Williams (1979, p.34) states that memory impairment is usually a *single* aspect of impaired functioning which is indicative of more *diffuse* neurological dysfunctioning. Memory impairment is frequently one of the primary identifiable symptoms of diffuse dysfunction and is sometimes even the sole symptom which is evident (Williams, 1979).

Failure to recall is both characteristic and a frequent initial manifestation of chronic and degenerative brain conditions (Lezak, 1983) such as Korsakoff's disease (Lezak, 1983; Lishman, 1978; Williams, 1991) and Alzheimer's disease (Lezak, 1983), and a hallmark of both deliriums and dementias (Williams, 1991). Further, Cooper, Sagar and Sullivan (1993) found that memory impairment in recently diagnosed Parkinson's disease patients suggested that the time disorientation deficits they experienced may be limited to short-term memory defects. Research shows that memory deficits are the most frequent cognitive impairment in adolescents and adults (Levin *et al.*, 1983; Oddy, 1984). Literature also reveals that traumatic head injuries are the prevalent cause of brain damage in individuals under 42 years of age (Lezak, 1983). Memory disorders are one of the most undeviating and differentiating symptoms of head trauma (Binder & Rattock, 1989; Brooks, 1989; Levin *et al.*, 1983; Lezak, 1983, 1989a, 1989b; Vakil, Arbell, Gozlan, Hoofien & Blachstein, 1992; Williams, 1979). Previously, individuals with apparently minor head injuries and no loss of consciousness were "written off as "neurotic" or victims of "compensationalitis", a malady that purportedly resolves with a cash consideration for the patient's "pain and suffering"" (Lezak, 1989a, p.xi). Yet, Lezak (1989a, p.xi) states that even these individuals regularly "suffer permanent residual cognitive deficits".

As well as being a general indicator of neurological damage, memory impairment is often the first presenting indication or an early sign of specific instances. For example, Diamond *et al.* (1992) found memory impairment in likely carriers of the genetically-transmitted Huntington's disease, prior to any identifiable symptoms. They concluded that immediate memory is sensitive to cognitive functioning and noted that their results concurred with similar studies. Further, Lyle and Gottesman's (1977) and Lyle and Quast's (1976) research with the same disease led them to suggest that memory defects occur prior to the overt manifestations of symptoms. In a study of intellectually superior older adults, Naugle, Cullum and Bigler (1990) found memory impairment before other cognitive functions of neurologically impaired adults revealed any decline. Memory dysfunctions also occur in HIV-positive individuals before symptoms appear, and increase significantly once they do (Law *et al.*, 1994; Mapou & Law, 1994).

The extent of memory impairment is also frequently used as a measure of damage incurred (Lezak, 1983; Levin *et al.*, 1983). For example, Huppert & Beardsall (1993) recorded that memory deficits in early dementia patients increased as neurological deficits did. Further, memory dysfunctions appear to be prognostic. In this regard, Vilkki, Ahola, Holst, Öhman, Servo & Heiskanen (1994) reported that the length of post-traumatic amnesia significantly differentiated between brain damaged individuals whose work status resumed normality a year after injury, and those who did not. The extent of memory impairment has also been useful in identifying the type of brain damage. For example, Vilkki *et al.*, (1992) found that patients with intercranial haematomata revealed greater memory impairment than other brain damaged patients. In addition, residual memory deficits often remain once the causative problem has been eliminated or other neurological functions have recovered (Gronwall, 1989; Levin *et al.*, 1983; Walsh, 1991; Yeates & Mortesen, 1994).

Further, memory impairment reveals neurological damage that even IQ scores do not. Oscar-Berman, Clancy & Weber, (1993) found that neurological impairment was indicated by the considerable discrepancies between scores on IQ and memory tests, yielding a convincing identification of Korsakoff patients. Memory impairment is so varied that it appears in individuals with conditions as diverse as children undergoing radiation for paediatric tumours (Ris & Noll, 1994), multiple sclerosis victims (McIntosh-Michaelis *et al.*, 1991) and epilepsy (Jambaqué, Dellatolas, Dulac, Ponsot & Signoret, 1993).

Clearly, the measurement of memory is an important factor to consider in neuropsychological assessment. A cautionary note is, however, necessary. Despite the significance of memory, like other cognitive functions it is not a fail-proof, single measure of impairment. For example, Dennis *et al.* (1991) warn that some locations of the brain which are invaded by tumours appear immune to memory deficits.

2.3.5 Types of memory in the perceptual processes

Like other aspects of neurological functioning, memory pertains to various perceptual functions. Reeves and Wedding (1994, p.3-4) differentiate between "declarative" or "fact" memory and "procedural" or "skill" memory. The former, as its name implies, involves the retention of facts which can be deliberately recollected and deals with information which is obtained through learning. This declarative memory is vulnerable to amnesic deficits. By contrast, procedural memory pertains to acquired skills or adjustable intellectual processes which are often not impaired in memory deficits (Reeves and Wedding, 1994). For example, Reeves and Wedding (1994) state that research reveals that individuals with memory dysfunction can learn diverse types of perceptual and motor skills although they are unable to recall the conditions under which these skills were gained. They add that the performance of neurologically impaired individuals on perceptual motor tasks can be close to that of non-impaired individuals while their performance on memory tasks is poor.

2.3.6 The duration of memory

2.3.6.1 Defining the parameters of immediate and short-term memory

Authorities vary on their definitions of immediate and short term memory (Lezak, 1983; Rahmann and Rahmann, 1991; Reeves and Wedding, 1994). However, for the purposes of this research it is sufficient to regard the immediate recall of the BG designs as a test of immediate memory and a second recall after a 10 minute delay as a delayed memory test which measures short-term memory.

2.3.6.2 Delayed memory

The importance of delayed memory is highlighted by a single case study of a head-injured subject, conducted by Markowitsch, Calabrese, Haupts, Durwen, Liess and Gehlen (1993). On the Digit Span Test and the copy version of the Ray-Osterreith Figure Test,[2] the subject obtained a score identical to the control subject and he performed *better* than the control on the immediate recall of the Ray-Osterreith Figure. It was only on the *delayed* version of this test that his neurological damage became apparent, as his performance was 30% lower than the control subject's performance.

Further research supports the sensitivity of a delayed recall test. Brook (1975, p.970) found that a delayed recall on the Benton Visual Retention Test produced an "impressive impact" and concluded that this adds a further "sensitive" and precise identification of normal and brain damaged. Gainotti and Marra (1994) also found that measures of delayed memory best differentiated between an Alzheimer's dementia and a pseudo-dementia resulting from depression. Research by Larrabee, Youngjohn, Sudilovsky and Crook III (1993) confirm this. Simon, Leach, Winocur and Moscovitch (1994) found that some measures of primary memory did not identify early Alzheimer sufferers while recall over brief delays did. Tröster *et al.* (1993) established that a delayed recall test differentiated significantly between victims of Alzheimer's and Huntington's disease where an immediate recall test did not. They concluded that sole dependence on immediate recall abilities would have "masked significant group differences" revealed by a delayed recall (Tröster *et al.*, 1993, p.783). Further, Lezak (1983) states that with visual memory, two recall tests assist in the

---

2. Like the BGT, the recall versions of this test involve memory for designs.

differentiation of specific features of memory impairments. She adds that the distortions of designs by neurologically impaired individuals on copy or immediate recall tasks are increased on a delayed recall version. If there are also visuospatial difficulties or perceptual disintegration problems, increased distortion or confusion of aspects of the design will occur (Lezak, 1983).

Memory abilities after an interference task are also relevant to the assessment of impaired functioning. For example, in certain stages of acute traumatic brain injury, memory is often impaired following a distraction (Gronwall, 1989). Further, Segalowitz, Unsal and Dywan (1992) established that head injured patients performed significantly worse than control subjects on a nonverbal (auditory) test of short-term recall which involved interference delays.

Wilkinson and Koestler (1983) note that in delayed memory tests, the likelihood of an item being recalled again increases if the item has been previously recalled and decreases if the item was not recalled in the prior trial. In contrast to the above, Youngjohn and Crook III (1993, p.457) revealed that well-educated, older individuals remembered more after a delay and they attributed this to the "consolidation" of memory. However, they add that their findings might not apply to lower educational levels.

2.3.7 The measurement of memory

Clearly, a memory test is indicated for the assessment of brain damage. Walsh (1987; 1991) maintains that a memory assessment should include the efficiency of immediate retention of information. In addition, Doubros and Mascarenhas (1969) caution that before perceptual-motor impairment is diagnosed the clinician should consider other facets of functioning, which include short term memory.

Swiercinsky (1979, p.240) notes that although Luria emphasizes the need for a memory test in a test battery, no single component on his battery solely measures memory, leading to the conclusion that "a separate test of memory functioning is not obviously present within this battery". Further, Purisch, Golden & Hammeke, (1978) found that the memory component of Luria's Neuropsychological Battery failed to discriminate significantly between organically impaired individuals and schizophrenics. Johnson, Mc Cusker and Boyd (1976) note that clinicians must often assess memory functions, to eliminate the possibility of severe impairment rather than to render a precise account of memory function. Due to time constraints this is not always done "rigorously with standardized memory instruments" (Johnson *et al.*, 1976, p.90). Clearly, a rapid, standardized measure of immediate and short-term memory is needed.

A suitable test for assessment of brain injured patients should be sufficiently uncomplicated for "confused" individuals to comprehend and rapid enough to be administered within the limited attention span of these patients. (Gronwall, 1989, p.23). It should also include a measurement of immediate and short-term memory (Gronwall, 1989). In choosing a memory test, the problems of assessment must be considered. For example, the problems with concentration and tiring easily exclude lengthy tests (Gronwall, 1989). Selection of material for visual memory testing is also complicated by the fact that minimal verbal interference should be involved (McFie, 1975).

2.3.8 Summary

The only verbal aspect of the BGT is the instructions. As already discussed (see 2.1.7: p.20), the BGT-*c* has a strong case law as an already sensitive test in diagnosing brain damage and differentiating between functional and organic impairment. However, as discussed above, the importance of memory as an indicator of brain damage is such that this facet of neurological functioning cannot be ignored. The expansion of the BGT by adding a BGRT-*i*, would enhance its diagnostic value since an additional component of neurological functioning, namely immediate memory, which is sensitive to diffuse neurological dysfunctions, would be measured. It is certainly more economical to use a BGRT if the BGT-*c* has already been used, than to administer a different memory test. Further, it is hypothesized that a BGRT-*d* would be even more sensitive to brain damage considering the literature reviewed above (see 2.3.6.2: p.43). Two BGRTs would build additional functions into a widely used, easily accessible, quick and easy-to-administer test.

2.4 The BGRTs

This fourth section reviews literature concerning the BGRTs. This section argues that the BGT-*c* can be enhanced by the addition of a BGRT-*i* and BGRT-*d*. The rationale and the functions of the BGRT are discussed. Then studies involving the BGRTs are reviewed, followed by the variations of the BGRT. The expansion of the BGT-*c* with a BGRT-*i* and a BGRT-*d* is proposed, and finally, the summary extrapolates the main issues regarding the BGRTs and the expanded BGT.

2.4.1 The rationale and the functions of the BGRT-*i*[3]

The BGRT-*i* was developed primarily to increase the test's ability to differentiate between organic brain damage and functional pathologies such as schizophrenia (Lezak, 1983). Lezak (1983, p.388) maintains that the BGRT-*i* is a good measure to investigate "visuospatial" memory. Garron and Cheifetz (1965, p.198) state that although the BGT-*c* and BGRT-*i* involve the same set of designs, the two versions do not measure "identical or even necessarily related functions". Their research outlines the function of the recall of BG designs in the following statement:

> Defects in memory for designs associated with organic brain pathology is a problem in its own right with a considerable literature to which the studies of BG drawing [*sic*] from memory may be more relevant (Garron & Cheifetz, 1965, p.198).

Such a memory for designs test was developed by Graham and Kendall and involves 15 geometric designs (Kendall, 1962). Shearn, Berry and Fitzgibbons (1974) used this test with psychiatric patients. The test revealed a significant difference on the test scores of patients whom clinicians had reason to believe may have possible neurological impairment and those for whom no presence of such damage was likely. Shearn *et al.* (1974) point out that most other studies which investigate the abilities of neuropsychological tests to identify brain damage are conducted on subjects known to have neurological impairment, which is often severe. Such information is of "limited relevance" in trying to confirm suspicions of mild brain damage in psychiatric patients whose impaired functioning is predominantly due to

---

3. The literature studies reviewed here do not specifically call the immediate recall of the BG designs the BGRT-*i*. However, they do comply with the definition of the BGRT-*i* (see 2.3.6.1: p.43). Therefore the terminology of BGRT-*i* will be appropriately used to avoid confusion with the BGRT-*d*.

their emotional state (Shearn *et al.*, 1974, p.1099). The sensitivity of the recall of designs in the detection of neurological impairment is therefore an asset when considering the use of the BGRTs to identify neurological impairment in psychiatric patients.

Although no literature regarding the functions of the BGRT-*d* could be found, it is proposed that the BGRT-*d* measures similar functions to that of a BGRT-*i*. However, it must also be noted that a difference in functions measured is likely to exist as literature already reviewed (see 2.3.6.2: p.43) reveals that delayed memory tests detect impairment even when immediate memory tests do not.

The combination of the BGT-*c* and BGRTs would result in two different measures of brain damage, namely perceptual motor skills and memory. This dual testing method may assist in increasing diagnostic accuracy. Both false negatives and false positives on either the BGT-*c* or the BGRTs could possibly be reduced by examining contradictions or accordances between the scores of the two tests. For example, the score of an individual which falls within the organic range on the BGT-*c*, but which indicates a good recall ability on the BGRT or *vice versa*, could alert the clinician to possible diagnostic errors. Alternatively, a BGT-*c* score indicating neurological impairment could be further supported by a very low BGRT score. Thus, the addition of the BGRTs would increase diagnostic accuracy in that they would be sensitive to another facet of brain damage not measured by the BGT-*c*.

In addition, it was discussed earlier (see 2.1.7: p.20) that while neurologically impaired individuals may copy designs accurately they would still be unable to execute drawing tasks accurately. The BGRT involves drawing rather than copying and therefore it could detect false negatives which might occur on the BGT-*c*. In addition, brain damaged subjects can perform similarly to normal individuals on perceptual motor tests (see 2.3.5: p.42). Further, Tolor (1958, p.17) observed from his research on the BGRT that "the immediate memory of psychogenic patients tends to be superior to that of organic patients". The above considerations support the hypothesis that the addition of two BGRTs will enhance the diagnostic abilities of the BGT-*c* both in the identification of brain damage and the differentiation between organic and functional impairment.

Research has shown that the BGRT is a more accurate measure of memory than some other neuropsychological tests. For example, Rogers and Swenson (1975, p.921) found that the BGRT measured a "more global sort of memory as reflected in long-term or short-term memory or a more general factor of intelligence". They concluded that it was an accurate screening measure of memory. Arbit and Zager (1978) investigated the functions of specific neurological tests. Their results indicated that the BGRT assesses memory ability better than the Graham-Kendall Memory-For-Designs Test does. The BGRT was also found to be "superior" to the Digit Span Test in its ability to differentiate between functional and organic impairment (Tolor, 1956, p.308). Schulberg and Tolor (1961, p.349) stated that the BGRT's role in ascertaining whether or not neurological damage is present could be of "empirical validity".

2.4.2 The frequency of use of the BGRTs

Despite its potential, the BGRT-*i* is less popular than the BGT-*c*. Schulberg and Tolor (1961, p.349) note with surprise that only 21% of clinicians surveyed "spend the few minutes necessary" to administer the BGRT-*i* frequently. This contrasts with the 77% of their respondents who used the BGT-*c* at least once a week. They attribute this lack of use to the fact that those who are "unsophisticated in the use of the recall phase reject this procedure on the basis of an alleged lack of validity" (Schulberg & Tolor, 1961, p.350). Although more current information regarding the status of the BGRT-*i* could not be found in published literature, indications are that these statistics still accurately reflect the current use of the BGRT-*i*. For example, in 1980, Rogers was only able to find fewer than six studies which investigated the BGRT-*i* in children. Only two studies of a BGRT-*d* version could be found.

While Lezak (1983) stated that the BGRT-*i* is valuable to assess visuospatial memory, she added that the lack of standardization of both test administration and scoring in the studies conducted has resulted in inadequacies which have not yet been resolved. Others also criticize the lack of standardization of the BGRT (Billingslea, 1963; Field *et al.*, 1982; Olin & Reznikoff, 1957).

In this section, it is argued that the failure by many clinicians to use the BGRT-*i* is attributable, at least in part, to three factors: (i) contradictory and/or negative research layout findings; (ii) the lack of a formalized system with which to score the BGRT-*i*; and (iii) the absence of any substantial normative data stratified for age and education. The first two factors are interrelated and mutually influencing, since without a standardized, efficient scoring system studies are likely to be flawed. Unsuccessful studies may in turn lead to decreased interest in the BGRT-*i* itself, which may result in fewer studies being conducted.

As with the BGT-*c*, the BGRT-*i* has been used to explore a variety of factors (e.g. Collaer & Evans, 1982; Edwards, 1980; Finch Jnr, Spirito, Garrison & Marshall, 1983; Goodstein, Spielberger, Williams & Dahlstrom, 1955; Peek & Olson, 1955; Rector, 1984; Sabatino & Becker, 1969; Sabatino & Ysseldyke, 1972; Sczechowicz & Hinrichsen, 1980; Shein, 1975; Smith & Winnick, 1979). In addition, a number of studies have used the BGRT-*i* to successfully differentiate between organic and functional impairment (e.g. Holland & Wadsworth, 1979; Levine & Feirstein, 1972; Lyle & Gottesman, 1977; Lyle & Quast, 1976; Tolor, 1956, 1958).

2.4.3 The ability of the BGRT to identify brain damage and to differentiate between functional and organic impairment

Tolor (1956) appears to be one of the first researchers to publish studies regarding the proficiency of the BGRT-*i* in differentiating between organic and functional impairment. This study consisted of three sample groups - organically impaired, schizophrenic and "compulsive" subjects (Tolor, 1956, p.308). The latter group was comprised of patients who experienced seizures, regardless of type or aetiology (Tolor, 1956). The BGRT-*i* was able to differentiate between the three groups at a 1% level of significance. The organic group recalled the least designs, the functional group the most, and the convulsive group scores fell between the two. Differences remained significant even when the scores were corrected for IQ. Tolor (1958) conducted a second study, replicating the first but imposing stricter scoring conditions. He found a significant difference between the recall abilities of schizophrenic and organic individuals, as well as between the organic group and a third group comprised of functional non-psychotic patients. Reznikoff and Olin (1957) found that stringent scoring of the BGRT-*i* yielded a difference, significant at the 5% level, between organic and schizophrenic patients. However, they commented that the overlap was too great for this method to be used in individual cases.

Holland and Wadsworth (1979) investigated the differential abilities of four versions of the BGT - namely the BGT-*c*, the BGRT-*i* , the BIP and a BIP-recall. They administered these tests in the order mentioned, to 40 inpatients, half of which had brain damage and half of which were schizophrenic. With IQ held constant, only the two recall versions were able to differentiate significantly between the two groups. Notably, it was found that the two recall scores were able to differentiate between the two groups "independently" of age, IQ, estimated premorbid intelligence, socioeconomic status and social abilities (Holland & Wadsworth 1979, p.126). Armstrong (1965) also found that the BGRT-*i* could differentiate between pathological aetiologies, where the BGT-*c* did not.

Levine and Feirstein (1972) compared the abilities of a number of neuropsychological tests to differentiate between organically impaired, schizophrenic and medical patients. The scores of the BGT-*c* and those of the BGRT-*i* differentiated significantly between the schizophrenic and organically impaired individuals. However, this was not true for the medical and schizophrenic group. The researchers conclude that:

> Interestingly, the simpler functions are as discriminating as are the more complex ... The very short and relatively simple BGT ... was found to be as effective a screening measure of organicity as was the lengthy Halstead-Reitan Battery (Levine & Feirstein, 1972, p.511).

Prediction of organic impairment is also important (Lezak, 1983). The BGRT-*i* has been successfully used in predictive studies. A rigorous longitudinal retrospective study by Lyle and Quast (1976) used the BGRT to evaluate individuals who had genetic links with Huntington's disease victims. This disease is a degenerative, hereditary condition, which includes cortical involvement.. They found significant differences between normal individuals, those who already had Huntington's disease and those who were in a pre-morbid state, i.e. those who manifest no symptoms at the time of testing, but who later developed the disease. However, Lyle and Quast (1976) stated that the accuracy of the BGRT, with hit rates of 67% and 68%, was not high enough to be used to predict individual cases. A second study also involving Huntington's disease patients and their blood relatives, repeated the investigation of the predictive abilities of the BGRT-*i* (Lyle & Gottesman, 1977). "Striking differences" were found between the scores of offspring of Huntington's victims who later developed the disease and the scores of those who did not (Lyle & Gottesman, 1977, p.1020). The BGT accurately predicted the true condition in 72% of the subjects. There were more true negatives (78%), than true positives (58%). Among the subjects classified as premordid, 39% developed overt symptoms of the disease within two years. However, the researchers cautioned that an innate but unavoidable flaw in the study is that some of the individuals in the group who showed no signs of Huntington's disease could still develop it at a later stage. Using the BGRT-*i* with children, Sczechowicz and Hinrichsen (1980) found that normal children recalled more designs on the BGRT than learning disabled children did.

2.4.4 Versions of the BGRT

2.4.4.1 The tachistoscopic version

This variation, involving a recall after a five-second exposure, is described by Wepman in personal correspondence with Lezak (Lezak, 1983) and is a mixture between BGT-*c* and a BGRT. It is normally administered first, followed by the BGT-*c* and then a BGRT. This author assumes that this version originates from Wertheimer (Wulf, 1922), as he experimented with a 30-second exposure, although no direct acknowledgment of this can be found. It is referred to as the 'Bender Gestalt Memory Test' by some researchers (Arbit & Zager, 1978; Becker & Sabatino, 1971; Sabatino & Becker, 1969). This can become confusing as other versions of BGRTs exist. Further, no research could be found which unequivocally demonstrates the tachistoscopic version to be a measure of memory. Thus adherence to the nomination of 'tachistoscopic' is preferable.

Little research regarding this method was found. Smith and Winnick's study (1979) is one of the few available on the subject. They administered the tachistoscopic version, the BGT-*c* and a BGRT-*i*, in this sequence (Smith & Winnick, 1979). They found that children who received two exposures to the BG designs, namely the tachistoscopic and BGT-*c*, prior to the BGRT had a higher recall rate than those having a single exposure. Thus, it appears that a practice effect exists for the BGRT with two exposures to the BG designs. Sabatino and Becker (1969) found a high correlation between the BGT-*c* and the tachistoscopic version. Variations of this version included an extended 40-second exposure to the designs rather than a five-second one (Goodstein *et al.*, 1955) while Becker and Sabatino (1971) and Sabatino and Becker (1969) allowed a five-second exposure and then implemented a 15-second delay before the designs could be drawn. It appears this was administered after the BGT-*c*.

The utility of the tachistoscopic version has not been well established by research. It demonstrated only a minimal increase in differentiation capacity between organic and non-organic patients, when compared with the BGT-*c* (Snortum, 1965). Smith and Winnick (1979) acknowledged that the diagnostic significance of this version has not been established. Lacks (1984) adds that there is inadequate proof of its accuracy. Further doubt is cast on the usefulness of this version by Bender's original theory (1938). She notes that the ability to perceive a gestalten within a limited time increases with age and experience. An inadequate exposure time to the BG designs prevents the design from being experienced in its entirety and as a result reproductions were inclined to revert to a more primitive configuration. This has possible negative implications for adults with low levels of education and for children, whose experience might be limited. Literature covering this version is limited and more research is needed before its diagnostic value can be ascertained. Moreover, the practice effect could confound a BGRT-*i*.

2.4.4.2 The BGRT-*i*

Most of the research conducted on the BGRT-*i* implements an immediate recall once the BGT-*c*, or in some cases, the tachistoscopic version, is completed and the designs have been removed (e.g. Anderson & Rallis, 1981; Armentrout, 1976; Goodstein *et al.* 1955; Holland & Wadsworth, 1979; Hutton, 1966; Levine & Feirstein, 1972; Lyle & Gottesman, 1977; Lyle & Quast, 1976; Olin & Reznikoff, 1957; Tolor, 1956, 1958).

There is little normative data available. In many instances scores are not given (e.g. Arbit & Zager, 1978; Clark, 1982; Collaer & Evans, 1982; Edwards, 1980; Peek & Olson, 1955; Shein, 1975). Further, some normative data is of no use as the studies reveal methodological flaws. For example, the study of Imm, Foster, Belter and Finch Jnr (1991), concerning recall in psychiatric inpatients, presents no diagnostic breakdown, making it difficult to establish for which type of population group their norms are applicable. Finch Jnr *et al.* (1983) made the same error on a sample of children referred for psychological evaluation because of learning and/or behavioural disorders. Additionally, the tabulated data of Finch Jnr *et al.* (1983) is statistically incomprehensible as it presents two different sets of contradictory data for the same age and IQ scores. The data reveals that no child under 8,75 years old has an IQ over 79 while all children 13 years and older have IQs of 110 or higher. Clearly such data is irrelevant here. Since BGRT scores will differ with the method of evaluation used, normative data relevant to this study will be presented later (see 9.5: 148) once the BGRTs' scoring method has been decided upon.

2.4.5 The BGRT

Only two studies on a BGRT-*d* could be traced in published literature (Edwards, 1980; Goodstein *et al.*, 1955). Edwards's study (1980) investigated practice effects and instructional sets. After rehearsal in drawing the designs, the subjects completed a BGRT-*i*. A BGRT-*d* was administered the following day. The second study, by Goodstein *et al.* (1955) involved the administration of a Draw-A-Person task after the administration of the tachistoscopic version of the BGT. Neither of the two studies investigated neurological functioning and no normative data is reported. Since they were the only published studies on BGRT-*d* which could be found, it appears that no normative data on such a version of the test are available.

2.5 Summary

The BGT-*c* is diagnostic in itself and it circumvents many of the problems experienced in assessment (see 2.1.8.1: p.22). For example, it is rapid to administer which limits the concentration difficulties and the fatigue experienced by many with brain damage. It is not vulnerable to malingering (see 2.2.1.4: p.31) and is ideal for evaluating psychiatric patients since it is able to differentiate between neurological and functional impairment with a high degree of accuracy. Research reviewed earlier indicates the value of immediate and delayed memory tests in neuropsychological assessment. The BGRTs will be more sensitive to mild brain damage, which presents a problem in assessment as discussed earlier (see 2.4.1: p.46). Defective memory is almost pathognomic in traumatic brain injury, even in cases of mild damage that are otherwise difficult to detect (see 2.3.4: p.40). Thus, the addition of a BGRT-*i* and a BGRT-*d* will make the BGT-*c* more sensitive to diffuse brain damage and thereby enhance the diagnostic efficacy of a test which is already widely used, easily accessible, and rapid to administer. It is therefore proposed that in this study, the BGT-*c* is expanded with a BGRT-*i* and a BGRT-*d*. This will be referred to as the expanded BGT.

As previously discussed (see 2.1.5: p.12), Lezak (1983) notes that the highest diagnostic accuracy is obtained by combining tests which detect specific impairments with those which measure a more diffuse dysfunction and those which elicit signs which are pathognomic of neurological impairment. The expanded BGT fulfils these three conditions. As already discussed (see 2.1.7: see p.20), the BGT-*c* is a measure of visual motor perception which can deteriorate when neurological impairment exists. The BGRTs are able to identify more diffuse impairment. Thus memory tests measure a more diffuse dysfunction, fulfilling the second of Lezak's (1983) requirements. In addition, inherent features within the expanded BGT scoring systems are regarded as organic signs, and these will be detailed later (see Chapter 7: p.115). The differential ability of the test is also enhanced by the BGRTs since it appears that in functional illnesses such as schizophrenia, recall abilities remain relatively intact (see 2.4.1: p.46).

However, this chapter reviewed the significant effects of education on the BGT-*c* and on memory (see 2.2.2.2: p.34), yet normative data for low-educated individuals is lacking. In an attempt to address this deficiency of data, it is proposed that there is a need to investigate the education effects on the BGT-*c* for low-educated adults. For the BGRTs to be of diagnostic value, the possible education effects must also be considered. It is therefore proposed that this research considers education effects on the expanded BGT and collects preliminary normative data for low-educated adults. Lacks (1984) has an educational category of less than 9 years of education. In order to facilitate cross-validation between the data in the present study and Lacks's (1984) data, the performance of adults with less than 9 years of education (i.e. an incomplete standard seven) will be investigated.

To pursue this type of research, certain problems concerning the use of the BGT, which have briefly been referred to so far, must be considered with a view to resolving them. The first issue involves which version of the BGT-*c* should be used. In the discussion of the variations of the BGT-*c* (see 2.1.6: p.16), only the BIP and Bender's original version have been submitted to rigorous research studies. Results with the BIP varied from excellent to dubious and more research on this version is needed. None of the versions discussed indicated an improvement in diagnostic accuracy. The original BGT-*c*, by contrast, has been researched since its inception in 1938 and therefore seems appropriate for the current research. The rapidity of administration is also important if the expanded BGT is to be of practical use (See Chapter 1: p.1). In this regard, the option of group testing does not interfere with the validity of the BGT-*c* and therefore appears appropriate to this research.

Other problems concern the lack of standardization in the administration and scoring of the expanded BGT which need to be addressed. In addition to existing administration problems, it is highly possible that certain criteria need to be considered in the administration of this test to low-educated adults. Also, the precise application of the group testing method needs to be considered, especially on the BGRTs which, according to published literature, has not been done before. The problems of the scoring of the expanded BGT are central to the research and must be resolved. These problems need to be considered in detail and attempts should be made to resolve them. Thus Part Two - Methodological Rationale for the expanded BGT has been included in this thesis.

PART TWO METHODOLOGICAL RATIONALE FOR THE EXPANDED BGT

Part Two attempts to resolve some of the problems surrounding the use of the BGT, and consists of four chapters. Chapter 3 considers the problem of the variety of methods used to score the BGT-*c* and an appropriate scoring system is decided upon. Then, Chapter 4 appraises the scoring systems used to evaluate the BGRT and suggests criteria for an appropriate BGRT scoring system. Chapter 5 considers certain unclarified issues concerning the administration of the test and concludes with an argument for the method of administration to be used in this research. Chapter 6 consists of a review of factors to be considered in the selection of subjects, and the criteria according to which subjects are selected are described.

For the expanded BGT to be diagnostically useful, certain problems need to be resolved as already discussed (see 2.2.3: p.36). Specifically, these involve the problems surrounding the lack of standardization of both the scoring and the administration of this test. The expansion of the BGT satisfies the first objective stated in Chapter 1. However, it was also stated (see Chapter 1: p.1) that for the expanded BGT to be useful in daily assessment situations, it must be "practical and economic to administer and score". The scoring of these tests is problematic, since three protocols for each subject must be evaluated. Scoring systems which are rapid to score, in addition to being accurate identifiers of neurological impairment, must be found to provide a solution to this problem. Therefore Chapter 3 will consider the scoring of the BGT-*c* while Chapter 4 will discuss the scoring of the BGRTs. As with other aspects of the BGT, the administration varies from one researcher to another, as already discussed (see 2.2.2.1: p.33). Because of this, specific administration-related issues are highlighted and decisions are made regarding the administration procedures to be followed.

CHAPTER 3: A CRITICAL EVALUATION OF METHODS USED TO INTERPRET THE BGT-*c*

In this section, the surfeit of scoring systems which have been used to evaluate the BGT-*c*, are noted[1]. Subsequently, the problems surrounding the multiplicity of scoring systems are reviewed and the counterproductiveness of this profusion of systems is discussed. Next, Bender's (1938) original method of interpretation is reviewed. This is followed by a critical evaluation of some of the major systems used in research literature. Finally, the Lacks (1984) system is critically considered and argument in favour of this system is advanced.

3.1 The problem of scoring the BGT-*c*

One of the most profound problems of the BGT-*c* is the proliferation of methods by which protocols are evaluated (e.g., Bender, 1948; Hain, 1964; Koppitz, 1975; Lacks, 1984; Pascal & Suttell, 1951; Pauker, 1976), as already discussed (see Chapter 1: p.1 and 2.2.1: p.30). Evaluation procedures range from a method which depends purely on subjective judgement (Bender, 1938; Dana *et al.*, 1983; Mosher & Smith, 1965; Wagner & Murray, 1969) through a number of systems which employ specific evaluating criteria (Hain, 1964; Koppitz, 1975; Lacks, 1984; Pascal & Suttell, 1951; Pauker, 1976), to the extreme parameters of the 312 weighted indices developed by Gobetz (1958, in Dana *et al.*, 1983). Thus numerous systems have emerged despite the fact that as far back as 1948, Billingslea (1963) called for standardization of the scoring system of the BGT-*c*.

Apart from the BGT-*c*, there appears to be no other neurological test for which such a large number of scoring systems exist. Whilst some of these systems have been especially useful in identifying brain damage and in discriminating between functional and neurological impairment, the very variety of scoring systems diffuses research efforts. It is evident that the results of studies are only meaningful and directly comparable with other research in which the same scoring system has been used. Researchers have mostly used either a single system, or three at the most, when comparing the abilities of different systems. Studies which each use a different scoring system are not conducive to productive research

---

1. In listing some of the minor systems, secondary references will be used. This information has been drawn from review papers.

as results cannot be compared nor validated when different measuring criteria are used. For example, the comparison of error scores of -13 on the Koppitz (1975) system, -34 on the Pascal-Suttell (1951) system and -5 on the Lacks (1984) system is meaningless. In addition to variations in the number of possible scoring errors on each system, the specifications which determine an error also differ across the scoring systems. This can be illustrated with a single example: Both Lacks (1984) and Koppitz (1975) consider the relative proportion of the designs as a scorable error. However, Lacks (1984) scores a part of a design which is one-third larger than another as incorrect, while Koppitz's system (1975) requires that the offending part of the design must be comparatively double or half the other part. In many cases, the criteria which determine an error do not overlap at all.

Another problem which arises is that many clinicians approaching the test for the first time are seldom aware of the wide choice of systems. Even if they are, it is unlikely that they have the time to evaluate all of these by weighing up the advantages and disadvantages of each system. While it makes some sense to use different scoring systems for the BGT-*c* when examining different functions (such as developmental maturity or reading ability), it seems neither rational nor practical to use different systems to evaluate a single criterion or related criteria such as the presence of brain damage and the differentiation between functional and organic impairment.

3.2 The proliferation of scoring systems for the BGT-*c*

A scoring system is quintessential to any neurological test. Lezak (1983, p.388) aptly states that many researchers have attempted to establish a "workable system" on the BGT-*c* in order to produce diagnostically-valuable scores. The emergence of the various scoring systems on the BGT-*c* and how they came to be developed is both complex and often insufficiently described. Detailing each one of these would necessitate an entire thesis in itself, and is obviously beyond the scope of this research. However, these systems will be briefly mentioned in chronological order in an attempt to indicate the need for a single, valid and rapid method of scoring the BGT-*c*. Scoring times of the various systems of the BGT-*c* vary from 20 seconds to a few hours, according to (Dana *et al.*, 1983). Hutt and Brisken (1960) note that some systems can take up to five hours to score. Field *et al.* (1982, p.838) note correctly that:

> In general, the empirical research literature with regard to the validity of the B-G scoring systems is difficult to interpret because of ... the use of inappropriate or poorly defined criteria.

Hutt's system, which he began using in 1945, was perhaps one of the earliest noted systems (Billingslea, 1963). It originally consisted of 27 scoring points which other researchers subsequently applied in part, by selecting only some of the 27 criteria (Harriman & Harriman, 1950 in Billingslea, 1963; Hanvik, 1953, in Billingslea 1963). Billingslea's own scoring system was published in 1948 (Billingslea, 1963) and involved 38 factors which used 137 indices to score. Billingslea (1963, p.237) admitted that his system was "too cumbersome for clinical research". Billingslea (1963) also refers to the scoring systems of Bell (1948), Buros (1949), Peek and Quast (1951) and Wooltmann (1950). Kitkay (1950, in Billingslea, 1963) used graph paper to establish the size of designs copied and to obtain data on 25 indices. In 1951, Pascal and Suttell (1951) published a scoring system which became one of the more popular ones and is still used today (e.g. Brilliant & Gynther, 1963; Carlson, 1966; Doubros & Mascarenhas, 1969; Haynes, 1970; Keogh & Smith, 1961; Korman & Blumberg, 1963; Landis *et al.*, 1974; Marsico & Wagner, 1990; Olin & Reznikoff, 1957; Swenson & Pascal,1953). Lesser known systems include that of Guertin (1952, in Dana *et al.*, 1983) which was based on factor analysis. Gobetz (1953, in Dana *et al.*, 1983) developed a system which had 312 indices. Keller (1955, in Billingslea 1963, p.235) developed an even more elaborate system which featured 114 signs, each of which is applied to each one of the nine BG designs. This involves 1026 scoring points in all. He stated that his system was both easy and rapid to score (Keller, 1955, in Billingslea 1963). Billingslea (1963) cites another system developed in Japan by Okino (1956), which had 121 indices.

Some systems were developed especially for children. Keogh and Smith (1961) developed a system to determine visuomotor ability and Quast (1961) identified specific deviations as being indicative of neurological damage in children. Watkins (1976, in Parsons & Weinberg, 1993) developed a system to identify learning problems in children. Koppitz (1975) established a system which has become both well known and widely used (e.g. Cerbus & Oziel, 1971; Finch Jnr *et al.*, 1983; Frey & Pinelli Jnr, 1991; Henderson *et al.*, 1969; Isaac, 1973; Kaspar & Lampel, 1972; Keogh & Smith, 1968; Marsh, 1972; Mc Connell, 1967; McNamara *et al.*, 1969; Nielson & Sapp, 1991; Sabatino & Becker, 1969; Simensen, 1974;

Snyder & Kalil, 1968; Taylor, Kauffman & Partenio, 1984; Thweatt, 1963; Welcher *et al.*, 1974). Her earlier work also included scoring items which related specifically to neurological damage and signs which indicated a high degree of emotion (Holroyd, 1966b). Koppitz's Developmental Scoring System (Koppitz, 1975) evaluates the degree of visuomotor perception in children relative to their maturity levels, which are gauged by chronological or mental age. Hence the term "developmental" in her nomination of the system.

Hain (1964) developed a system to differentiate between functional and neurological impairment. In 1966, Canter introduced the BIP (see 2.1.6.3: p.18) and modified the Pascal-Suttell (1951) system to score this new version. Parsons and Weinberg (1993) mention a method by Jansky and de Hirsch (1972) based on quality of production (Obrzut *et al.*, 1972). The Cooper-Barnes system tried to eliminate most of the subjective elements of judgement required by most scoring systems (Cooper & Barnes, 1966; Cooper *et al.*, 1967). This system compares the different proportions of designs and entails intricate measurements in millimetres, the calculation of ratios, the subtraction of absolute differences and finally the summation of these differences. It has a high interscorer reliability and successfully differentiated neurologically impaired subjects from a mixed sample of organic and psychiatric and normal subjects. But the procedure is tedious as, for example, the scoring of design A requires six different measurements in millimetres, in addition to all the related mathematical operations, which themselves must be considerably time-consuming.

Contemporary systems include Pauker's system (Pauker, 1976) and Lacks's (1984) adaptation of the Hutt-Briskin 12 discriminating signs of brain damage. In addition, a Qualitative Scoring System using only six designs was published by Brannigan and Brunner in 1989 (Brannigan & Brunner, 1991). The Sugar Scoring System, referred to by Parsons and Weinberg (1993), was unpublished when they used it, and also used only six designs. Hilgert (1985) developed a method using part of Hutt's scoring system and chose the indicators which most facilitated computer processing. Field *et al.* (1982) created a scoring system of their own, in a sense, by combining *some* elements of each of the scoring systems of Hutt, Pascal-Suttell and Pauker. Another factor which complicates issues surrounding scoring systems of the BGT-*c* is that some researchers, for example Hutt and Briskin (1960) have developed more than a single system or have continually modified their systems so that several versions of the same systems exist.

Despite the large number of scoring systems, or perhaps because of this fact, many of these have made only brief appearances in research literature and then disappeared. i.e., they have been only used in a few studies by other researchers or have not been used at all. Although at least 25 systems have been mentioned in the above review, it is not a fully comprehensive review on all the scoring systems' published literature. No doubt there are further variations which have been produced over the years. Perhaps one reason for the diversity of systems which have been derived from this single test can be attributed to the fact that Bender's (1938) original version of the test was non-scored. This is discussed in the next section. The better-known scoring systems, those of Pascal-Suttell (1951), Hain (1964) and Koppitz (1975), are then discussed in detail. Next, the Hutt-Briskin (1960) system is mentioned briefly. These methods are then evaluated with regards to their suitability for the current research. An evaluation of Lacks's (1984) adaptation of the Hutt-Briskin system follows.

3.3 A critical evaluation of the better-known methods of interpreting BGT-*c*

3.3.1 Subjective evaluation

Bender (1938) relied on clinical judgement to evaluate the BGT-*c*. Clinical judgement involves making a decision as to whether or not a protocol indicates neurological impairment, by considering the protocol and forming an opinion. Thus no scoring method is used. Research concerning the accuracy of clinical judgement is contradictory. It has differentiated accurately between functional and organic pathology with an 86,6% precision (Wagner & Murray, 1969). However, there are two problems with this study. Firstly, the five evaluators participating in this study were highly experienced in evaluating the BGT-*c* subjectively. Dr Laurette Bender herself was part of the team and this level of experience is not representative of average clinicians who have to interpret the BGT-*c*. Two of the evaluators were qualified at Ph.D level and the other two held M.A. degrees. Secondly, this study used a matched-pair research design which casts doubt on the validity of the results obtained.

A matched-pair design involves pairing two cases which represent the different levels of the independent variable. In this study, brain damaged individuals were paired with non-brain damaged individuals. The researcher then needs only to identify which individual belongs to which level. However, in real assessment situations, such pairing does not exist and this type of design is not possible to replicate. The success of a research study which cannot be applied practically must be questioned. Eno and Deichmann (1980, p.42) state that while the matched-pair design best eliminates nuisance variables it is "one of the most artificial and least representative of the prediction problems clinicians face".

Other research documenting the use of subjective judgement has been less successful. Bowland and Deabler (1956) used this method and found it differentiated between normal, neurotic, schizophrenic and organic subjects at a level better than chance. However, the overall diagnostic accuracy was just below 50%. Weinstein and Johnson (1964) also used a matched-pair design, pairing schizophrenic patients with individuals suffering from epilepsy with uncontrolled seizures. The protocols were matched in three ways: protocols from epileptic patients before and after undergoing a temporal lobectomy were matched with each other; pre-operative epileptic patients' protocols were matched with the schizophrenic group; and the post-operative epileptic protocols were matched with the schizophrenic group. The accuracy of the subjective judgement of clinical psychologists was above the chance level at 65,4% while the lay judges had an overall accuracy of 62,9%. While these percentages were better than chance, the accuracy of differentiation was not statistically significant. In their study, Weinstein and Johnson (1964) conclude that there is an anomaly between the confidence clinicians have in the BGT-*c*'s ability to identify brain damage and the actual results achieved. They add that the patients in the study, i.e. hospitalized schizophrenics and temporal-lobotomized organic patients, present more extreme cases than those a clinician normally encounters in a referral population. They predicted that in a referral population, accuracy would be even lower. Additionally, the accuracy of this study is further weakened by the use of a matched-pair design, which has already been discussed above. Although they conceded that the BGT-*c* must, in some way, have the ability to differentiate between organic and non-organic individuals, they stated that this ability is insufficient to diagnose individual cases and suggested its use as a differentiating tool be discontinued. This study places a question mark on the efficacy of subjective judgement.

The degree of experience necessary to interpret the BGT-*c* using subjective judgement is illustrated by Goldberg's (1959) study. He compared subjective judgement with the Pascal-Suttell scoring system. The subjective scorers ranged from unskilled individuals to staff psychologists. While both systems had a diagnostic accuracy greater than chance, Goldberg (1959, p.33) concluded that the results "suggest that chances for misdiagnosis could be increased by utilizing the Bender Gestalt Test".

As an adjunct to the study, Max Hutt, who developed his own scoring system, and who is a leading expert on the BGT-*c*, spent 20 hours examining the 30 protocols. He achieved a diagnostic accuracy rate of 83% - nearly 20% higher than the other scorers. Goldberg (1959) concluded that trained clinicians were no more accurate in applying subjective judgement to the BGT-*c* than untrained individuals:

> unless, perhaps, ... [their] training includes years of intensive work with the instrument in question. If he is not a real expert in the use of the Bender, a clinician will find that his secretary can probably do this particular job of differential diagnosis as well as himself (Goldberg, 1959, p.32).

Apart from contradictory findings in the accuracy of subjective judgement, this method has a number of other problems. Firstly, while the level of expertise required is appropriate to Bender's knowledge of the BGT-*c* , it is not practical in the modern clinical context. The years of experience necessary to become an expert evaluator using this method necessitates a degree of specialization which is impractical, if not impossible, for most clinicians. When considered in context, the BGT-*c* is one of many assessment tools, and assessment is usually only one of a number of skills demanded of clinicians. In addition, few clinical training courses include interpretation of BGT-*c* (Rogers, 1980). Secondly, subjective judgement is time-consuming. Even an expert such as Dr Max Hutt took nearly 40 minutes to evaluate a single protocol in order to obtain a high diagnostic accuracy. Thirdly, since subjective judgement cannot be quantified, it can also not be standardized, and therefore objectivity cannot be established. This renders research unsuitable for methodological documentation. It also means that cross-study reliability cannot be established nor can inter-study comparisons be made. Additionally, normative data cannot be collected on this method. Clearly this manner of evaluation of BGT-*c* protocols is neither quick nor standardized.

3.3.2 The Pascal-Suttell scoring system

Pascal and Suttell (1951) used a system of 105 weighted scorable error deviations. The system was developed primarily to differentiate between normal and psychiatric individuals. Scoring criteria for the first BG design (design A) are omitted without any reason provided. Pascal and Suttell (1951, p.6) state that their scoring rationale is based on an attempt to differentiate between "psychogenic" patients and non-patients. They admit that the distortions they score as errors could be attributed to damage in the cortex *or* impaired ability attributable to functional dysfunctions. Thus no factor in their system is designed to differentiate between neurological or emotional impairment. Since they found both age and education effects on the BGT-*c*, their normative data derived from more than 200 protocols is applicable to a 15-year-old to 50-year-old age group with at least 9 years of education. The Pascal-Suttell (1951) scoring system is one of the most widely used systems, mostly on the adult population for which it was intended (e.g. Brilliant & Gynther, 1963; Butler *et al.*, 1976; Canter, 1966; Carlson, 1966; Doubros & Mascarenhas, 1969; Haynes, 1970; Keogh & Smith, 1961; Korman & Blumberg, 1963; Landis *et al.*, 1974; Marsico & Wagner, 1990; Olin & Reznikoff, 1957; Swenson & Pascal, 1953).

However, research is contradictory. Some researchers claim a high accuracy rate in the identification of brain damage or the differentiation between the brain damage and functional impairment (e.g. Brilliant & Gynther, 1963; Korman & Blumberg, 1963; Landis *et al.*, 1974; Marsico & Wagner, 1990; Song & Song, 1969; Swenson & Pascal, 1953; Wagner & Marsico, 1991), while others have produced disappointing results (Bowland & Deabler, 1956; Butler *et al.*, 1976: Goldberg, 1959; Mosher & Smith, 1965; Shaw & Cruickshank, 1956; Snortum, 1965; Weinstein & Johnson, 1964). Only some of the unsuccessful studies apply the system to an age younger than that for which it was intended (Gilbert & Levee, 1967; Shaw & Cruickshank, 1956; Song & Song, 1969). Thus, other reasons for the unsuccessful studies must exist.

Despite its popularity, the Pascal-Suttell scoring system has a number of problems. Firstly, the manner in which the scoring system was developed has resulted in a significant overlap between functionally impaired and organically impaired individuals (Pascal & Suttell, 1951). This is clearly detrimental to accurate differentiation as some research shows (Lezak, 1983; Mosher & Smith, 1965), yet differentiation is of primary importance in assessment (see

2.1.8.2: p.23). The authors (Pascal & Suttell, 1951) admit to a low reliability coefficient and the frequent occurrence of false negatives.

Bowland and Deabler (1956) found that this system was able to differentiate between normal individuals and patients who were organically impaired, neurotic and schizophrenic, at a level that was higher than chance. But they concluded that it was insufficiently accurate to diagnose individual cases. Supporting this view, Goldberg (1959) found that while this system differentiated between organic and psychiatric patients at a level better than chance, the low accuracy meant that the use of the BGT-*c* could actually increase diagnostic errors. Snortum (1965) used this scoring method to evaluate both a BGT-*c* and a tachistoscopic version. No significant differences on the BGT-*c* were found between hospitalized patients with diagnoses of 'alcoholic' and 'neurotic' or between the alcoholic group and the group of organically impaired individuals. The tachistoscopic phase differentiated between the control subjects, the alcoholic patients and the organically impaired patients. However, the neurotic patients did not differ from the control group and differed from the organic group at only a 0,2% level of significance.

Secondly, the large number of items on the scoring system and the fact that is weighted, makes it both complicated to learn and time-consuming to score. Lacks (1984) estimates that it takes a competent scorer about 20 minutes to score each protocol. The third problem is the limited age range to which this scoring method is applicable. Assessment and diagnosis are often important in children when schooling begins, or in older adults, after the commencement of neurological deterioration due to aging. Yet this system is not applicable to these populations. Further, Chouinard and Braun (1993) note that the expanding numbers of older people will result in the need for increased neuropsychological evaluations of elderly people. Fourthly, this system is limited to subjects with at least 9 years of education (Pascal & Suttell, 1951). A high percentage of the South African population of adults is poorly educated and worldwide, about one-third of the population is illiterate (Steinberg, 1990).

Finally, an additional negative factor of this scoring system, for the purposes of this research, is that it was not specifically developed to identify organic impairment, but rather cortical malfunction whether of a structural or behavioural nature. The authors' statements clearly indicate this fact in the two inserts below:

> 'Does the psychological record show signs indicative of damage to the cortex?'.... The answer ... involves the difficult and plaguing task of attempting to differentiate between deviations in performance due to psychogenic and histogenic, genogenic or chemogenic factors. The B-G test cannot, in the absence of other data, answer that question except occasionally in extreme cases which are also clinically apparent... Our data for cortical deficit and the differential diagnosis of psychogenic disorders has not been gathered systematically (Pascal & Suttell 1951, p.40).
>
> Performance on the B-G test can indicate damage to the cortex only when the damage shows its effect by pronounced disturbance of the ability to execute the test (Pascal & Suttell 1951, p.62).

Considering the above, this scoring system is not suitable for the current research. Despite certain inadequacies, the Pascal-Suttell scoring system (1951) has made a valuable contribution to the study of the BGT. Firstly, it has been highly successful at times (e.g., Landis *et al.*, 1974). Secondly, it has stimulated research and has also been used as a base for other scoring systems. Olin and Reznikoff (1957) used it as a basis for a BGRT scoring system, while Canter (1966) used it to develop a system to score his BIP Test. The most well-known adaptation of the Pascal-Suttell (1951) system is the Koppitz Developmental Scoring System (Koppitz, 1975). Koppitz modified this system initially to identify learning problems in children (Thweatt, 1963). It has become one of the most used and successful methods to evaluate the BGT-*c*.

3.3.3 The Hain (1964) scoring system

Hain (1964) developed a scoring system with the specific intent of being able to differentiate between functionally and organically impaired individuals. He established the frequency of specific errors in the protocols of organically and functionally impaired and normal subjects. He then used specific, mutually exclusive distortions from these protocols to create a 31-point weighted diagnostic scale. He reduced these criteria to 15 points of discrimination, with weighted scores varying from 1 - 4. However, on cross-validation, his scoring system misidentified 41% of the brain damaged patients. He also admitted to frequent false negatives. Other studies have also found Hain's (1964) system to be deficient in various ways, including the fact that it increases the risk of diagnostic errors, particularly in cases of mild impairment (Butler *et al.*, 1976; Lacks & Newport, 1980; Mosher & Smith, 1965). For example, while Mosher and Smith (1965) found Hain's system differentiated between brain damaged and normal subjects at a rate better than chance, errors were so numerous that the system was regarded as inadequate for individual evaluation. Using Hain's suggested 'cut-off' score, more than 76% of the brain damaged individuals in the study were misidentified. Mosher and Smith (1965) suggest that the BGT-*c*, when evaluated according to Hain's method, actually increased the occurrence of diagnostic errors in a sample with a low incidence of brain damage. Butler *et al.* (1976) found that the system did not differentiate between neurologically and functionally impaired patients. They added that this system was of minimal or no use in borderline cases.

This system appears to have inconsistent results and a low level of accuracy. Another deterrent to using this system is, as Lacks (1984) notes, that it has never been published. Certainly, no normative data could be found on this scoring system. Considering the above, Hain's (1964) scoring system is inappropriate to the research in question.

3.3.4 The Koppitz (1975) Developmental Scoring System

Koppitz's (1975) Developmental Scoring System has become a particularly favoured gauge of developmental visuomotor perception in children (e.g. Cerbus & Oziel, 1971; Finch Jnr *et al.*, 1983; Frey & Pinelli Jnr, 1991; Henderson *et al.*, 1969; Isaac, 1973; Kaspar & Lampel, 1972; Keogh & Smith, 1968; Marsh, 1972; Mc Connell, 1967; Mc Manis *et al.*, 1978; McNamara *et al.*, 1969; Nielson & Sapp, 1991; Sabatino & Becker, 1969; Simensen, 1974; Snyder & Kalil, 1968; Taylor *et al.*, 1984; Thweatt, 1963; Welcher *et al.*, 1974). It has a high validity and reliability as well as a high interscorer reliability which has been supported by numerous research findings (e.g. Kaspar & Lampel, 1972; Miller, Loewenfeld, Linder & Turner, 1963; Morsbach, Del Priori & Furnell, 1975; Obrzut *et al.*, 1972; Ryckman *et al.*, 1972). Furr (1970, p.78) refers to this as "one of the most useful scoring systems" for the BGT-*c*.

Koppitz (1975) operationalized the BGT-*c* with an error-scored, developmentally-orientated scoring system to gauge the level of visuomotor perceptual maturation in young children. She clearly states that like all developmentally-based tests, the score obtained using this method is only significant *in relation* to the subject's age. She states that children's visuomotor perceptual skills are sufficiently developed by the age of 10 for a child to copy the designs accurately when evaluated according to her system. Koppitz (1975) therefore concluded that her system was applicable to mature 5-year-olds through to 10 years 11 months, while above average children may reach the ceiling at 8 years of age (Koppitz, 1975). This is a point many researchers miss and as a result some report contradictory findings (e.g. Aikman, Belter & Finch Jnr, 1992; Andert *et al.*, 1978; Cerbus & Oziel, 1971; Condell, 1963; Culbertson & Gunn, 1966; Dierks & Cushna; Hutton, 1966; Finch Jnr *et al.*, 1983; Skeen *et al.*, 1982; Viljoen *et al.*, 1994; Yousefi, Shahim, Razavieh, Mehryar, Hosseini & Alborzi, 1992). Taylor *et al.* (1984) suggest an even lower age ceiling of 8 years old to which this scoring system may be appropriately applied.

Many researchers have also attempted to use this scoring system to investigate the presence of brain damage (Andert *et al.*, 1978; Arnold *et al.*, 1978; Condell, 1963; Eno & Deichmann, 1980; Friedman *et al.*, 1967; Mc Connell, 1967; Mc Manis *et al.*, 1978; Welcher *et al.*, 1974). However, Koppitz (1975, p.14) strongly refuted this as incorrect, since her system is "validated against general overall achievement" and scoring points not directly associated with visual-motor perceptual development were eradicated. She notes that her own and others' studies (Koppitz, 1975) have identified cases of brain impairment, but states that these are not diagnostically definitive. Instead Koppitz (1975) attributes this to the detection of a maturation lag which can also occur in normal children. Yet earlier work done by Koppitz (1962, p.542) included applying her system, which at that time she referred to as "The Bender Scoring System for young children", to identify brain damage in children. Even then, she noted that neurologically impaired older children with a high intelligence could execute good BGT-*c* protocols. She also notes in the preface of the second edition of her book (Koppitz, 1975, p.xiii) that "in some instances it was necessary to modify my earlier positions". The ability of her scoring system to detect brain damage is clearly one such alteration.

There are two reasons why this system is problematic for the research in question. Firstly, it is a developmental gauge and is not designed to identify brain damage. Secondly, the age range to which it can be applied is not suitable for adults.

3.3.5 The Hutt-Briskin (1960) scoring system

Hutt began using a scoring system to evaluate the BGT-*c* as early as 1945 (Billingslea, 1963). This consisted of 27 factors. In 1960, Hutt and Briskin proposed 12 differentiators of brain impairment, or 11 according to Lezak (1983, p.16). However, little detail of scoring or empirical validation was provided and these two researchers then turned their attention to more projective interpretations of the test in question.

Lacks (1984) correctly states that Hutt altered his criteria a few times. This is evident in research literature. In their book, *The Clinical Use of the Revised Bender-Gestalt Test*, Hutt and Briskin (1960, p.89) propose 12 "essential discriminators" of "intercranial damage". In addition, these authors present numbers "specific test factors and their interpretation" (Hutt & Briskin, 1960, p.47). They further present 'essential discriminators' for conditions as varied as psychoneurosis, schizophrenia, mental retardation and manic states (p.91). Each set of 'essential discriminators' also has an accompanying list of "associated discriminators" (Hutt & Briskin, 1960, pp.88-92). Further details of various systems and modifications are beyond the scope of this research, which focuses on the 12 'essential discriminators' of brain damage (Hutt, 1960).

3.4 An evaluation of the above methods

All the evaluation methods detailed here are marred by at least one of the following factors: (i) inherent validity and reliability problems; (ii) the lack of operationalized, replicable procedures; (iii) the difficulty of learning to apply the scoring system compounded by the lack of training at clinical levels; (iv) the lengthy time needed for evaluation which poses problems for busy clinicians; (v) the applicability of the scoring system to a very narrow age group or educational level; (vi) the lack of age-stratified normative data; (vii) the fact that the method has not been specifically designed to identify the presence of brain damage; and (viii) the deviation from the original concept of the gestalt, which is inherent to the test (Butler *et al.*, 1976; Dana *et al.*, 1983; Hain, 1964; Koppitz, 1975; Lacks, 1984; Lezak, 1983; Mosher & Smith, 1965; Pascal & Suttell, 1951; Wagner & Murray, 1969). Thus, despite the number of systems, there still appeared to be a lack of what Lezak (1983, p.388) calls a "workable system", at the time her book was published.

Mosher and Smith (1965) warn that if the BGT-*c* continues to be utilized in the differentiation of organic and functional impairment, an improved scoring system must be developed, or alternatively, meticulous training in subjective evaluation is essential. Despite such scoring problems, research shows that the test is still widely used and has an established value. Therefore, if a simple, accurate, standardized method of scoring could be used, its value would be enhanced.

3.5 The Lacks (1984) adaptation of the Hutt-Briskin (1960) scoring system

3.5.1 Lacks's (1984) operationalization of the 12 Hutt-Briskin (1960) discriminators of brain damage

Russell (1976, p.355) states that "apart from one study" which he cites as that by Lacks *et al.* (1970), the accuracy of the BGT-*c* in discriminating between schizophrenics and neurologically impaired individuals is not significantly higher than other neuropsychological measures. This study used Lacks's (1984) adaptation of the 12 Hutt-Briskin 'essential discriminators' for brain damage.

Lacks's (1984) adaptation of the 12 Hutt-Briskin discriminators of brain damage contains 12 error-scored criteria which apply to some or all of the designs. Lacks (1984) established a neurological cut-off point of -5 and scores of -5 or more are thus regarded as indicative of brain damage. This system is contained in full in Appendix 2, and is reproduced with the permission of Wiley Publishers (see Appendix 9B: p.265). Lacks (1984) includes nine observational cues as part of her scoring system (see Appendix 3: p.257). This system has the potential to be the "workable system" (Lezak, 1983, p.388) which many have tried to develop. She defined the criteria with more rigour and precision than Hutt and Briskin (1960), with the specific intention of developing a scoring system which is sensitive to organic impairment.

This system meets the criteria discussed above, which other systems have failed to do. Firstly, its differentiating ability has been successfully researched since 1963, with some studies attaining a diagnostic accuracy of 82% or higher when discriminating between organically and functionally impaired individuals (Brilliant & Gynther, 1963; Johnson *et al.*, 1971; Lacks, 1984; Lacks *et al.*, 1970; Lacks & Storandt, 1982; Marsico & Wagner, 1990).

The overall diagnostic accuracy of the system is 77% (Lacks, 1984). Lacks and Newport (1980) investigated four scoring systems including Lacks's (1984) adaptation of the Hutt-Briskin system. It produced an accuracy of 84% and was the most successful differentiator between organic and non-organic psychiatric inpatients presenting with organic, psychotic, personality disorder and chronic alcoholic symptoms. Further, Brilliant and Gynther (1963) reported an 82% accuracy using this system on a sample of 120 organically and neurologically disordered patients. This study found the BGT-*c* to be the best differentiating neuropsychological test of three - namely the BGT-*c*, the Graham-Kendall Memory-for-Designs Test and the Benton Visual Retention Test.

A study by Marsico and Wagner (1990) compared the Pascal-Suttell scoring system and Lacks's scoring system. They evaluated the protocols of 104 outpatients who were either organically impaired or suffered from functional psychological disorders. The Pascal-Suttell method produced an overall diagnostic accuracy of 73% compared to an accuracy on the Lacks system of 71%. In both cases, there were more false negatives than positives. However, the researchers point out that the Pascal-Suttell system had a organic "hit rate" of 50% while the Lacks system was slightly higher. Marsico and Wagner (1990) concluded that the systems were analogous in terms of their relative accuracy. Thus, the accuracy of the system has also been established by researchers other than Lacks.

The second advantage of Lacks's (1980) scoring system is that it is fully operationalized and is available in a published form, obviating the need to rely on the availability of training courses. Thirdly, a competent knowledge of the system can be gained in three to four hours of concentrated study (Lacks, 1984). Thus, learning the system is comparatively quick and easy. Research indicates the ease of mastering this system. Lacks and Newport (1980, p.352) used three different levels of scorers, namely "expert", "typical" and "novice" scorers. The highest level of diagnostic accuracy with was obtained by the novice scorer, i.e., an individual who had recently completed a B.A. (Psychology) degree, with no previous knowledge of tests for organic impairment. In addition, an inter-scorer reliability of 0,87 to 0,90 was reported (Lacks & Newport, 1980).

Fourthly, Lacks (1984) maintains that the system takes less than three minutes to score. This author has found this to be correct, once familiarity with the criteria is established. This is particularly relevant with the introduction of BGRTs (see Chapter 1: p.1), since the number of protocols trebles.

With regard to the fifth criticism of other scoring systems, namely the limited age applicability, Lacks's system (1984) has no restrictions in this area. This system has been tested on adults ranging from 17 to 87 years of age (Brilliant & Gynther, 1963; Johnson *et al.*, 1971; Lacks, 1984; Lacks *et al.*, 1970; Lacks & Storandt, 1982; Marsico & Wagner, 1990). Although no published literature of Lacks's (1984) system being applied to children exists, the present author's previous research applied it to children and obtained meaningful data for children ranging from 7,5 to 16 years of age (Dyall, 1993). It appears that this system may well be applicable to a younger population. It seems logical then, that it will also be successful in the evaluation of adults with low education.

Sixthly, a comprehensive bank of normative data on this scoring system exists. Lacks (1984) provides normative data which are stratified for age and education using large numbers. Her data also includes sample numbers norms for psychiatric and organic patients. The seventh advantage is that the system was specifically designed to identify brain damage.

Finally, the Lacks (1984) system remains true to the spirit of the original gestalt concept as scoring criteria are largely based on errors which deviate from the concept of the gestalt. For example, the error criterion of fragmentation specifies that an error point is scored if the gestalt is broken. In keeping with the principle of the gestalt, Lacks (1984) considers a protocol from an *overall* view, i.e. if rotation exists in one or in six designs, one error point, that of rotation - is noted.

3.5.2 A critical evaluation of the Lacks (1984) system

One of the aims of this research is to expand the BGT into a test which is more sensitive to brain damage. However, to be of practical use, it must be possible to score it rapidly with a scoring system that is simple and objective, in order to attain the highest validity and reliability. In the light of the discussion above, Lacks's (1984) adaptation of the Hutt-Briskin scoring method is the most appropriate one to use. Advantages of this method include: (i) its high diagnostic accuracy rate; (ii) the accessibility of the fully operationalized scoring system presented by Lacks (1984); (iii) its availability and the brief training period and ease of mastering this system; (iv) the rapid scoring time; (v) the broad age range to which it can be applied; (vi) the existence of normative data; (vii) the fact that it is specifically designed to detect brain damage; and (viii) its adherence to the concept of the gestalt.

Two studies criticizing the Lacks (1984) system have been found in published literature. The first, conducted by Margolis *et al.* (1989) reports that the BGT-*c* only accurately diagnosed 36% of geriatric patients. However, Hart (1993, p.185) calls this "misleading" and points out that these researchers failed to apply age-stratified norms to their subject sample which consisted of older adults. In the second one, Friedt and Gouvier (1989) state that organically impaired individuals were misdiagnosed as normal in nearly 50% of the cases. However, there is still a 74,1% accuracy in the differentiation between functionally and organically impaired subjects. Of the 'misdiagnosed' subjects, 15% of the functionally impaired subjects indicated organic impairment. This is not necessarily a misdiagnosis, as Friedt and Gouvier (1989) make no mention of any efforts taken to ensure that those individuals classified as functionally impaired had no possibility of brain damage. The authors also do not note whether the schizophrenic patients were taking medication, which could also affect results. Further criticisms of this study include the fact that all subjects were convicted criminals. Yet the control group came from this same population, which is not necessarily normal, merely because the control subjects revealed an absence of schizophrenia or organic impairment. Only control subjects with a neurotic or affective diagnosis or a history of drug abuse were excluded. Another factor is that the interscorer reliability in this study was only 78% - more than 10% lower than that recorded by Lacks (1984) as stated above. Also, the organic individuals in their study were older, had a lower IQ and education which can be significant in confounding results (see 2.2.2.2: p.34 and 6.3: p.107). For example, two subjects from the control group and one from the schizophrenic group who were "misclassified" as organic, were admitted by the researchers to have "subnormal" intelligence (Friedt & Gouvier, 1989, p.459).

Another factor which is discussed in detail later (see 5.2.3: p.98) is the importance of observation on the BGT-*c*, particularly when using the Lacks's (1984) method. Apparently this was not done as protocols were taken from subject files and the researchers note that they failed to score Lacks's (1984) optional item of excess time taken simply because this was not recorded on most of the protocols. A failure to adhere to the stipulations of a scoring system does not mean that the system is deficient. Finally, serious doubt is cast on the validity of this study not only because of the low number of subjects (10 in each group) but the fact that the researchers refer to 30 protocols, but evaluate the percentage of diagnoses which are correct, out of 40.

In contrast to the criticism of Lacks's (1984) system, it is worth noting, that in another study which seriously criticized the use of the BGT-*c* (Bigler & Ehrfurth, 1981), three protocols were replicated in the paper. According to the system used by Bigler and Ehrfurth (1981) which the researchers did not identify, they claimed none of the three protocols indicated brain damage, which had been confirmed by other methods. However, two of these scored within Lacks's (1984) organic range when marked by this author, using Lacks's (1984) system.

Although the Lacks system (1984) has a high degree of accuracy, it does appear to flounder at times with a degree of false negatives. But this is not unique and as already discussed (see 2.1.5: p.12), no neuropsychological tests are 100% accurate, nor are there any medical tests which produce total accuracy. Thus, a high degree of accuracy must suffice. However, the addition of the BGRTs could increase the accuracy of the BGT-*c* since memory impairment plays a prominent role in neurological dysfunction. In addition, the combination of functions measured by the expanded BGT strengthens the diagnostic accuracy of the BGT-*c* and the BGRTs. The problem with measures of diffuse functioning when used alone, is that psychiatric patients are inclined to render false positives as already stated (see 2.1.5: p.12). Lacks's (1984) system, as discussed above is more prone to produce false negatives (see also 3.5.1: p.70). Thus the combination of the BGT-*c* and the BGRTs supplement the weaknesses of each measure.

CHAPTER 4: A CRITICAL EVALUATION OF METHODS USED TO INTERPRET THE BGRTs

In this chapter, specific errors in scoring the BGRT[1] are detailed and then criticism of the BGRT, keeping these errors in mind, is presented. Subsequently, Weiss's (1970) BGRT system is reviewed and evaluated. Next, the attributes of a scoring system which would be suitable for the BGRTs proposed in this study, are formulated. Finally, it is suggested that a combination of the Lacks (1984) BGT-*c* scoring system and Weiss's (1970) BGRT scoring system could meet the criteria for a suitable scoring system for the BGRTs.

4.1 The need for a BGRT scoring system

Despite the proliferation of existing systems on the BGT-*c*, the BGRT by contrast does not appear to have a single operationalized system that has stood up to successive research studies. In most studies, scoring methodology on the BGRT is, at the very least, haphazard. A number of the studies which claim that the BGRT is inadequate have used scoring systems which are faulty in some way (e.g. Clark, 1982; Rogers, 1980; Sabatino & Ysseldyke, 1972; Shein, 1975). Considering the confusion surrounding the scoring of the BGRT, contradictions in research are understandable. Olin and Reznikoff (1957) indicate that many of the inter-study discrepancies regarding the BGRT may be significantly related to the measurements involved. They emphasized the necessity for the scoring of the BGRT to be standardized in order for comparisons to be made between studies conducted by different researchers. Lezak (1983) stated that while the BGRT is valuable, the lack of standardization of scoring in the studies is problematic. Other researchers agree (Dana *et al.*, 1983; Field *et al.*, 1982). The absence of standardization of any BGRT system makes comparisons between studies difficult and of dubious validity.

---

1. Only two studies could be found which use any form of a BGRT-*d*. Since neither of these provide details of their scoring systems, all literature cited in this chapter refers to studies conducted on the BGRT-*i*. Further, a scoring system applicable to the BGRT-*i* will also be applicable to a BGRT-*d*. Therefore, in this chapter, no distinction between the versions of the BGRTs will be made unless it is of fundamental importance.

4.2 Errors in scoring

Problems in scoring result from a number of errors which have been made, and include:

(i) the use of the Koppitz scoring system to evaluate the BGRT

(ii) the lack of detail regarding the scoring method used

(iii) insufficient details or oversimplification of scoring criteria

(iv) scoring which credits designs despite the fragmentation of the gestalt

(v) the acceptance of rotations as correct

(vi) the lack of a BGRT scoring system specifically designed to identify neurological damage

In addition, many researchers do not compare performance on the BGT-*c* with that of the BGRT. While this is not a scoring error as such, it reflects the loss of potentially valuable diagnostic information, which has already been collected, since the administration of the BGT-*c* is a prerequisite for the administration of the BGRTs.

4.2.1 The use of the Koppitz (1975) Developmental Scoring System to evaluate the BGRT

A number of studies using the Koppitz (1975) system on the BGRT lack validity. Firstly, the Koppitz Developmental Scoring System is based on developmental criteria. Koppitz (1975) clearly stated that her system is not intended as a detector of organic damage (see 3.3.4: p.67), and all but developmental factors were eliminated from the scoring design. Therefore, scoring the BGRT with this system means the recalled designs are being evaluated with criteria which indicate the presence or absence of developmental errors only. Despite this, researchers use this scoring system to evaluate the BGRT in an attempt to differentiate between different types of neurological status (e.g. Clark, 1982; Sabatino & Becker, 1969; Sabatino & Ysseldyke, 1972; Shein, 1975). For example, Shein (1975) found that the BGRT did not differentiate between brain damaged and normal children using Koppitz's system. Instead of claiming, like some, that the BGRT was worthless, she aptly concludes by calling for the establishment of a scoring system for the BGRT which is sensitive to neurological impairment in children. In addition, as with the BGT-*c*, some studies go beyond the appropriate age range allowed by Koppitz (Hutton, 1966; Sczechowicz & Hinrichsen, 1980).

Secondly, the Koppitz (1975) scoring system is a weighted error-scored method with negative points for errors. Although this is not a formalized BGRT scoring system, researchers who have applied it to the BGRT fail to state whether a design which is not recalled is given a score of 0 or the maximum error points for that incorrect design. Mathematically only a score which awards maximum error points is correct (D. Dizdar, personal communication, December 1994). Awarding a design which is not recalled a score of 0 is mathematically inaccurate on an error-scored system. For example, by hypothesizing that each design can score a maximum of two error points (i.e. -2), the failure to recall a single design *or* the recollection of one to nine designs without error, would both result in a score of 0. Reproduction from memory of designs with a total of three errors would have a score of -3. Therefore the recall of nine designs with three errors would produce a poorer score than recalling no designs at all.

Thirdly, by using this system, one score can denote more than a single condition, even if the failure to recall a design is correctly scored at the maximum error score (which in this hypothetical case is -2). Four designs recalled without error would thus result in a score of -10 (the number of designs not recalled multiplied by -2). However, the same score of -10 would be awarded to a protocol in which seven designs were recalled, of which one has no errors, and six have one error each. Thus -10 would signify four designs recalled without error *and* seven designs recalled with six errors. In addition, the Koppitz Developmental Scoring System is weighted, which means different types of errors can score different scores and not each design has the same number of scoring points. Thus the BGRT score is neither equivalent nor proportional to the number of designs recalled. In the assessment of memory it is arguable that it is preferable to have a system in which a single scoring level denotes one mutually exclusive condition only.

Research findings illustrate the problems associated with using this system to evaluate memory. Rogers (1980, p.862) used this method and concluded that the BGRT was "doubtful" and "inconsistent". Clark (1982) found that the BGRT did not differentiate between brain damaged and normal children. Sabatino and Becker (1969) and Sabatino and Ysseldyke (1972) also found that results on the BGRT were disappointing when using the Koppitz Developmental Scoring System to evaluate protocols.

Some research is further confused by researchers stating that they used the Koppitz Developmental Scoring system, but then referring to a *positive* number of designs recalled (Hutton, 1966; Smith & Winnick, 1979). For example, Hutton reports 1,8 designs recalled. This score is unlikely to be a Koppitz score even if one assumes the "-" sign was omitted, as this would mean that all his subjects scored better when they recalled the designs than they did when they were copying the designs. Presumably he used some other criteria to evaluate the BGRT which he failed to describe.

Sczechowicz and Hinrichsen (1980) used the Koppitz system to evaluate the BGRT in a slightly different way. They first established which designs had no Koppitz errors (apart from designs 7 and 8 which they allowed two errors). They then counted those designs without errors (or with no more than two, for designs 7 and 8), to establish the number of designs correctly reproduced. While this is a mathematically acceptable scoring method, two problems remain. Firstly, it is still *developmental* processes which are being evaluated, while memory tests are most often used to investigate the presence of *neurological* impairment. Secondly, the arbitrary decision of allowing designs 7 and 8 to have two errors which presumes these are more difficult to remember, weights the system in an unscientific manner. Conversely, others claim that designs 3 and 4 are the most difficult to recall (Armstrong, 1965; Goodstein *et al.*, 1955; Olin & Reznikoff, 1957; Reznikoff & Olin, 1957; Tolor, 1956; Tolor 1958). In addition, Tolor (1958) found that designs 7 and 8 are amongst the easiest to recall and Reznikoff and Olin (1957) and Tolor (1956) supported the view that design 8 is the easiest to recall.

4.2.2 Failure to report the scoring methodology used

It is impossible to evaluate the scoring systems used on a number of studies since no mention is made of which scoring criteria were used to score the BGRT (Arbit & Zager, 1978; Becker & Sabatino, 1971; Collaer & Evans, 1982; Edwards, 1980; Hutton, 1966; Levine & Feirstein, 1972). This might be an acceptable error on a test which has only a single, widely accepted scoring system. However, this is a serious omission when the criteria for evaluation are different in almost every published study.

4.2.3 Inadequately defined and over-simplified scoring criteria

A large proportion of the BGRT studies reviewed have scoring criteria which are either inadequately defined or over-simplified or both. While the nine BG designs initially appear very simple, scrutinizing a series of protocols reveals how many different versions of a single design can be drawn. Reznikoff and Olin (1957, p.184) highlight the scoring dilemma succinctly by stating that "the problem of finding suitable criteria for measuring the recall immediately presented itself when the data were first examined". The broad range of possible permutations in the reproduction of these designs is endorsed by some of the scoring systems which have been developed. For example, Keller's system (1955, in Billingslea, 1963) has a total of 1026 possible errors in the reproduction of the nine designs. This illustrates that the scoring criteria for evaluating the nine designs needs to be clearly and precisely defined if a BGRT system is to be standardized and replicated.

However, such precision does not exist in most of the studies reviewed. Since it would be laborious to cite all the criteria used, some examples will illustrate the point. For example, Peek and Olson (1955, p.186) use the term "completely recalled" when referring to an entire design and "a clearly recognizable element" when referring to a design partially recalled. "Clearly recognisable" is a term which is also used by Rogers (1980, p.860). Studies by both Finch Jnr *et al.* (1983, p.88) and Imm *et al.* (1991, p.442), which had some researchers in common, scored according to criteria which established that designs had "no major distortions, omissions or variations". Even *et al.* (1988, p.990) used the term "a fair resemblance" to "the basic gestalt of the original" as a definition of his scoring procedure. Armentrout (1976, p.832-833) scored designs "depending on degree of completeness" and accredited designs with a full point when "all major components were present with no distortions other than minor ones (such as incorrect no. of dots)". The term "correctly reproduced" used by Lyle and Quast (1976, p.230) is even more vague. A later study (Lyle & Gottesman, 1977, p.1013) uses a scoring criterion of a design being "drawn correctly". Tolor (1956, p.305) simply states that "whole and part figures were counted". Tolor (1958, p.15) had even fewer parameters, stating that the recall score constituted the sum of "whole and part, correct and incorrect" designs.

It is clear that none of these studies could obtain equivalent results if they were replicated by other researchers. When criteria are insufficiently described, subjective assessment becomes a factor. What one scorer perceives as 'correct' or 'a major distortion' will differ from the opinion of other scorers. Clearly more definitive criteria are needed for an effective scoring system if the BGRT is to be sufficiently standardized to allow for the replication of research.

4.2.4 Breaking the concept of the gestalt

In their attempts to score the BGRT, many researchers have lost sight of the fundamental tenet of the test, namely the organization of the gestalt (see 2.1.2: p.6). This point cannot be sufficiently emphasized. The basic tenet of gestalt theory and therefore that of the BGT, is that a healthy perceptual process involves primacy, i.e., the dominance of the whole gestalt over the individual parts which make up the total pattern. For example, Bender (1965, p.190) states that the test will expose impairment in adults which result in, among other things, "disorganisation of perceptual motor gestalten". Thus, awarding part of a design a scoring point, is incongruent with the rationale of the BGT. Yet many have made this error (e.g., Armentrout, 1976; Lyle & Quast, 1976; Peek & Olson, 1955; Rogers, 1980; Rogers & Swenson, 1975; Tolor, 1958; Smith & Winnick, 1979). Smith and Winnick (1979, p.440) gave full credit for "partially reproduced" designs. Peek and Olson (1955, p.187) awarded a half score to part of a design recalled when it was "recognisable". Rogers (1980) went as far as fully crediting a design which contained a mixture of elements from more than one design. Tolor (1958, p.15) included designs which were incomplete as well as ones which were "incorrect".

It should be noted that the predictive accuracy in identifying premorbid sufferers of Huntington's disease, using the BGRT, increased from one study to another when the scoring procedure was changed. In the first study, Lyle and Quast (1976) credited partially recalled designs with half scores. In the second study (Lyle & Gottesman, 1977), a design was scored in keeping with the gestalt principle - i.e., a full point was awarded to a design accurately reproduced. Surprisingly, the significance of these scoring alterations and concomitant increase in diagnostic accuracy was not noted by the authors.

Two other errors, apparent in some scoring systems which are also contrary to Bender's (1938) fundamental rationale of the gestalt are: (i) allowing a lack of integration and (ii) permitting perseveration. Both of these errors reveal disorganization in patterning, the significance of which has already been discussed (see 2.1.2: p.6). Thus, scoring these errors as correct once again breaks the tenet of the gestalt. Nevertheless, some researchers have scored such errors as correct (Anderson & Rallis, 1981: Rogers & Swenson, 1975). Further, perseveration is regarded as a strong organic sign, as already discussed (see 2.1.5: p.12).

4.2.5 Scoring rotations as correct

Lezak (1983, p.149) discusses neurological signs and states that there are some such signs which are "isolated response deviations that, in themselves, may indicate the presence of an organic defect". She cites research which indicates that the rotation of designs is one such 'neurological sign'. Another factor which indicates the importance of rotations is the Minnesota Percepto-Diagnostic Test in which some of the BG designs are evaluated on rotations only (see 2.1.6.1: p.16). This test is successful in identifying neurological impairment (Lezak, 1983).

Rotation of the BG designs has been a subject of much research (Black, 1973, Field *et al.*, 1982; Griffith & Taylor, 1960; Hanvik, 1951; Jernigan, 1967; Shein, 1975; Smith & Martin, 1967; Snyder & Kalil, 1968; Weiss, 1971a), and has been found to be a significant factor both in differentiating between normal and cerebrally impaired individuals and between those who are functionally or organically impaired (Griffith & Taylor, 1960; Jernigan, 1967). In addition, Smith and Martin (1967) found that the inability to correct rotations with learning cues and practice effects was a significant sign of neurological deficit (Smith & Martin, 1967). Many scoring systems incorporate rotations as a scorable error (Haynes, 1970; Koppitz, 1975; Lacks, 1984; Pascal & Suttell, 1951; Weiss, 1970). Hanvik (1951, p.194) concluded from extensive experience of using the BGT with children that:

> the rotation of the Bender figure is an even more malignant sign when it occurs in children than when it occurs in adults and that rotation is a visual-motor aberration almost pathognomic for brain damage in children.

He reported that electroencephalogram abnormalities were evident in 80% of children who rotated at least one BG design. Despite this, researchers have erred by scoring rotations as correct (e.g. Even *et al.*, 1988; Finch Jnr *et al.*, 1983; Imm *et al.*, 1991; Rogers & Swenson, 1975) or partially correct (Armentrout, 1976).

4.2.6 The lack of a BGRT scoring system designed to identify brain damage

On the BGRT there appears to be a total lack of an operationalized scoring system which has been specifically designed to detect neurological impairment. The BGT-*c* is administered for a variety of reasons, not all of which involve the measure of neurological impairment (see 2.1.4: p.10). But the memory test is presumably administered specifically to detect neurological impairment as memory dysfunctions are one of the most undeviating symptoms of neurological damage (see 2.3.2: p.38). This makes the lack of such a system more serious. For example, the system of Olin and Reznikoff (1957) is based on the Pascal-Suttell (1951) scoring system, which is not designed to identify brain damage (see 3.3.2: p.64).

4.2.7 Comparison of the BGT-*c* and BGRT scores

Many researchers make no effort to compare their BGT-*c* and BGRT scores (e.g. Even *et al.*, 1988; Finch Jnr *et al.*, Goodstein *et al.*, 1955; 1983; Imm *et al.*, 1991; Lyle & Quast, 1976; Peek and Olson, 1955; Tolor, 1956, 1958), while only Goodstein *et al.* (1955) state that this omission was intentional. In addition, these researchers do not even present the scores of the BGT-*c* at all. It appears that they have used the BGRT *instead* of the BGT-*c* rather than as an additional test to increase diagnostic accuracy. It makes sense to maximize the tests to their full diagnostic potential by obtaining as much data as possible. Since the BGT-*c* has to be executed in order for the BGRT to be administered, information from the BGT-*c* might as well be used.

Olin and Reznikoff (1957) took note of this problem. They state that any scoring system employed for the BGRT should also be appropriate to the BGT-*c* in order to facilitate a direct comparison between the two tests. They add that relatively few studies exist which try to operationalize the BGRT. Olin and Reznikoff (1957) attempted to quantify the BGRT with the same scoring system as the BGT-*c*. They modified the Pascal-Suttell negatively weighted scoring system into 117 negatively weighted indices. High correlations were found between the BGT-*c* and BGRT, with a good diagnostic accuracy. The problem with Olin and Reznikoff's (1957) attempts is that the time needed to learn both scoring systems and to score two protocols per subject (or three in the case of this research) makes it too time consuming to be of any practical value to busy clinicians. Additionally, the Pascal-Suttell system was not designed to identify brain damage (see 3.3.2: p.64) while this is the purpose of most memory tests.

The author hypothesizes that the scoring of protocols using cumbersome and time-consuming methods is a deterrent to using information from the BGT-*c*. However, Lacks's (1984) scoring system takes less than three minutes per protocol to score and resolves this problem. It is also necessary to use compatible scoring methods for the BGT-*c* and the BGRT(s) to make a comparison between the tests possible and meaningful. With such a system, it could be expected that a high correlation between the copy and recall phases will occur.

4.3 Addressing criticism of the BGRT in light of the above

The most severe criticism of the BGRT has been mooted by Rogers (1980). He called the BGRT "unreliable" because it has a very high intra-item variability whilst test variability and reliability were extremely low (Rogers, 1980, p.859). In addition, he stated that investigating the quality of the recalled designs "are of little utility and can only be expected to yield inconsistent results" (Rogers, 1980, p.862). He adds that the BGRT is "of doubtful utility ... and further research is not likely to be productive" (Rogers, 1980, p.859). However, Rogers (1980) based his condemnation of the BGRT on inexact conclusions. Firstly, on the BGT-*c*, he used subjects older than the age which Koppitz stipulated (see 3.3.4: p.67). Secondly, it appears that he used two methods to score the BGRT, although he does not explicitly state this. The first method was Koppitz's (1975) method which is unsound unless a design not recalled was awarded the maximum possible error score (see 4.2.1: p.76), yet this is not stated. The second system scored designs as recalled provided that they were "recognizable ... even if parts of two or more figures were mixed ... in one drawing" (Rogers, 1980, p.860). This approach is totally contrary to gestalt principles (see 4.2.4: p.80). Further, comprehensive studies of research indicate that a lenient scoring system is not a good measure on the BGRT, and this will be discussed in detail later (see 4.5.2.1: p.86).

4.4 A critical evaluation of Weiss's (1970) scoring criteria

Weiss (1970) obtained age-stratified normative data on the BGRT-*i* for children. However, his study has several limitations. For example, he makes no the mention of controlling for normality in his subjects. Also he does not say whether the administration method involved group or individual testing nor does he provide details of his administration proccedure. Further, Weiss (1970) does not give BGT-*c* scores corresponding to the BGRT scores he presents. Weiss's (1970, p.143) scoring system stipulates that BG designs must be recalled "without or no more than one single slight distortion". Based on this stipulation, Weiss (1970, p.143) presents his criteria as follows:

> Following this distinction, we excluded, in the research under report, the following types of distortions;
>
> a. Simplifications, i.e. the replacement of a complex (part) figure, e.g. hexagon or square by a diamond or triangle respectively; 'swallowing-up' or 'cutting-off' of a corner in Figs 7 or 8 or shortening of Figs. 1 or 2 were not excluded.
>
> b. 'Extreme' i.e. edge-to-edge perseverations in contrast to 'moderate' ones, where the number of units (dots or triplets) exceeds that of the design, but does not reach 'extreme' proportions.
>
> c. Rotations for more than 45°.

Despite the above criticism, Weiss's (1970) scoring system is the one which most closely attains the criteria of being clearly defined and practical to use in terms of time. However, his operational criteria, although better defined than most others, are still inadequate. Therefore, Weiss's (1970, p.143) definition of allowing no more than one 'slight distortion' and his concept of perseveration need to be further operationalized. Further definitions of minor distortions allowed were not provided while in the actual reproductions minor variations are prolific, as already discussed (see 4.2.3: p.79). More definite defining criteria are needed to ensure scorer consistency and, at a later stage, inter-scorer reliability and validity.

4.5 The need for a BGRT scoring system and its necessary criteria

Considering the literature presented above, the need for an appropriate scoring system for the BGRT remains. It is hypothesized that the BGRT(s) would be very useful if correctly scored. However, in light of the review of the BGRT scoring systems, there is clearly a need for an operationalized BGRT scoring system which is rapid, efficient and specifically designed to detect brain damage.

Further, no scoring system can be diagnostically useful if no normative data which are stratified for education are available. This is so because, according to literature already reviewed (see 2.2.2.2: p.34), education exerts a significant influence on neurological test scores and specifically on the scores of the BGT-*c*, and on memory tests.

As already discussed (see 4.1: p.75), both Olin and Reznikoff (1957) and Lezak (1983) maintain that the absence of a standardized scoring system on the BGRT poses problems. It must be borne in mind at all times that the purpose of introducing the BGRTs in this study is to measure a different facet of neurological impairment other than that which is already detectable on the BGT-*c*. As Garron & Cheifetz (1965) note (see 2.4.1: p.46), the BGRT is an indication of memory. Before considering the actual components of a BGRT scoring system, it is necessary to establish the rationale and criteria for such a system.

4.5.1 Considering past errors in scoring systems

Any attempt made to develop a BGRT system should not repeat past errors. Thus, errors made by researchers should be noted. By extracting information from the evaluation of scoring systems used to measure performance on both the BGT-*c* and BGRT (see 3.3: p.61 and 4.2: p.76), it is suggested that the following criteria are necessary for a BGRT scoring system:

(i) The scoring criteria used should be defined adequately and in sufficient detail to allow the replication of the scoring system by other researchers

(ii) The scoring system should retain the concept of the gestalt and penalize errors that break the gestalt. Thus, integration of designs should be deemed necessary, while fragmentation should be penalized.

(iii) Rotations should be regarded as errors as they are neurological signs.

(iv) A BGRT scoring system should be rapid to score and should not require a lengthy training period.

(v) Scores obtained using a BGRT scoring system should be reported in a clear, precise manner

(vi) As the identification of brain damage is the purpose for which the BGRT is being used and the focus of this research, a scoring system used to evaluate the BGRT should be based on factors designed to detect neurological impairment.

(vii) Further, it would be useful if the BGT-*c* and BGRT(s) could be compared. Any BGRT scoring system developed would be more useful if it was compatible with the BGT-*c* which would allow for the comparison of the tests.

4.5.2 Other relevant factors pertaining to a recall scoring system

In addition to the above factors, two other parameters need to be established. These are: (i) the degree of stringency of a recall scoring system; and (ii) the mathematical principle to be used in a scoring system

4.5.2.1 The degree of stringency

The degree of stringency used in the evaluation of recall designs must be considered. Rogers and Swenson (1975) compared two methods of scoring the BGRT - a stringent and a more lenient measure. They concluded that it is of "little consequence" how rigidly one scores the recall protocol (Rogers & Swenson, 1975, p.921). Some researchers have followed their example and opted for lenient scoring (e.g., Finch Jnr *et al.*, 1983; Imm *et al.*, 1991). In a later study, Rogers (1980) scored mixed designs as correct and found the data to be unreliable. Yet, from an evaluation of research findings, it appears that more stringent as opposed to more lenient scoring systems have been successful. For example, the study of Sczechowicz and Hinrichsen (1980) found that when the designs were considered to be adequately recalled, provided they had "sufficient resemblance" to the original designs, no significant differences were found between normal and learning disabled children (Sczechowicz & Hinrichsen, 1980, p.466). However, they also applied an alternate system in which only designs which scored no Koppitz error-points were considered correctly recalled (some leeway was allowed for designs 7 and 8, for which a maximum of two Koppitz errors were allowed in order to be scored as accurately recalled). This manner of scoring resulted in significant differentiation between the sample groups.

Other studies also indicate that a more stringent scoring criteria on the BGRT improves diagnostic accuracy. As already discussed (see 4.2.4: p.80), the study of Lyle and Gottesman (1977) differentiated between neurologically impaired and normal individuals with greater accuracy than the one by Lyle and Quast (1976) did. The second study used more stringent scoring criteria. This is further supported by Olin and Reznikoff (1957), who investigated three levels of stringency in the evaluation of BGRT scores. Only the most stringent of three sets of scoring criteria were able to differentiate between organic and functional impairment. Thus, a more stringent evaluation of the BGRT appears to result in a greater discriminatory ability.

In light of the above therefore, it was decided that a BGRT scoring system should be comprised of relatively stringent as opposed to lenient measures of evaluation. A stricter measure of evaluation will also provide greater reliability and inter-scorer correlation. However, it must also be borne in mind that this is a recall task and therefore, some allowance must be allowed for human error. To expect absolutely perfect reproductions is unrealistic, particularly if the scoring system is going to be applied to young and low-educated samples.

4.5.2.2 The mathematical principle used in a scoring system

There are three possible mathematical models on which a recall scoring system can be based: (i) an error-scored recall system, (ii) a combination of error-scored designs and the sum of designs recalled, and (iii) a system that first establishes the criteria for a recall design and then reports the number of designs that meet these criteria.

(i) An error-scored recall system

Scores on the BGT-*c* are mostly error scores. For each deviation from the original design which is specified as an error, a error point is scored (e.g. Koppitz, 1975; Lacks, 1984; Pascal & Suttell 1951) and some systems weight these errors (e.g. Koppitz, 1975; Pascal & Suttell, 1951). When using a weighted system, more than one error point can be scored for a single error, depending on the significance and degree of that particular error.

Recall is, in itself, a positive concept, i.e., a design is either correctly recalled, incorrectly recalled or it is not recalled at all. To score a recall test with an error score, while mathematically acceptable, is complicated and difficult to assess. For example, in a spelling test consisting of 10 words, it is clear that 10/10 means that there are no words misspelt, 9/10 means that one word is misspelt and so on. However, when error scoring is used, one score can represent more than a single condition as already discussed (see 4.2.1: p.76). It is obviously preferable for a recall scoring system to have each score denoting a single condition only in order for a clinician to assess the memory capacity of the patient. However, the problem of the multi-significance of a single score would once again be present.

(ii) A combination of error-scored designs and the sum of designs recalled

The second option of scoring methods is one with which each recalled design is awarded a positive mark (or marks) and is then scored for errors. This method would mean that the positive and negative scores would cancel each other out to an extent. Again, the problem of the multi-significance of a single score would emerge. Hypothetically, if each recalled design is worth a single point, nine designs recalled would result in a score of +9. If the designs contained six scorable errors, these would score -6 points making the total score of 3. However, a score of 3 could also be obtained by the recall of three perfect designs or by the recall of six designs with three mistakes.

(iii) A positive recall scoring system

A positive scoring system is simpler and easier to interpret than either of the above methods. The duplicity of the meaning of a single score could then be avoided. This would involve each design being scored either as correct (or partially correct) or as totally incorrect or absent. In a BGRT with 9 designs, the total score would be nine, or multiples of nine if more than one point is given for each design. If each correct design is awarded a single mark, the highest possible score would be 9 and the lowest score would be 0. Each score would then denote a single, equivalent situation. Thus, 6 would mean six designs are adequately recalled, and only this. It is suggested that this is the most appropriate method of scoring a memory test.

Although some researchers have used half points to credit partially correct designs (Armentrout, 1976; Lyle & Quast, 1976; Peek & Olson, 1955; Rogers & Swenson, 1975; Smith & Winnick, 1979; Tolor, 1958), this is contrary to the concept of the gestalt as already discussed (see 4.2.4: p.80). It is therefore suggested that no designs should be awarded partial marks. It is further suggested that each of the designs should have the value of a single mark.

All that remains to be established is which criteria should eliminate a design as inadequately recalled, which will be discussed in the next section.

4.5.3 Criteria which will establish the correctness of a design as adequately recalled

In deciding on the criteria which will establish the correctness of a recalled design, the nature of a memory test must be considered and compared to that of the BGT-*c*. For example, Lacks's (1984) system has an organic 'cut-off' point of -5 errors. Therefore, one, two or even four errors are not jeopardizing, and the BGT-*c* involves *copying* the designs. With the BGRT, however, a single error can render a design as inadequately *recalled*. Considering this and the discussion on stringency (see 4.5.2.1: p.86), a balance should be attained. This balance should allow for a realistic degree of stringency, while at the same time allowing for the fact the designs are being recalled, not copied.

4.5.4 The compatibility of BGT-*c* and BGRT scoring systems

In the existent research, little comparison has been made between the BGT-*c* and the BGRT (see 4.2: p.76 and 4.2.7: p.82). From the discussion in these sections it seems irrational to administer the BGT-*c* which is clinically valuable and then not utilize the information obtained. However, for a comparison to be valuable, a degree of correlation between the two scoring systems must exist. This is possible if the scoring systems are based on similar evaluating criteria. This could be accomplished by using similar criteria to establish if a copied design should be given a negative error point or whether a recalled design is sufficiently remembered to score a positive point.

4.5.5 Combining Lacks's (1984) BGT-*c* scoring system with Weiss's (1970) BGRT scoring system

Weiss's (1970) BGRT scoring system was considered the best one available (see 4.4: p.83). It meets many of the above criteria in that it does not break the gestalt nor does it allow rotations. The significance of the preservation of the gestalt and of scoring rotational errors has already been emphasized (see 2.1.2: p.6; 4.2.4: p.80; 4.2.5: p.81). It is also quick to score which has also been stressed as a necessary criterion (see 4.5: p.84). However, it was criticized for not defining criteria sufficiently and stated that more definite criteria are needed (see 4.4: p.83). On the other hand, Lacks's system (1984), which is being used in this research to score the BGT-*c* is clearly operationalized. Furthermore, it is a system which is specifically designed to identify brain damage. It was decided to use Lacks's (1984) criteria for the BGT-*c* as clear cut-off points within the confines of Weiss' system, to supplement criteria insufficiently defined by Weiss (1970). Where Weiss's (1970) criteria are vague, Lacks's (1984) system will be used in most instances to clarify them. This is facilitated by the fact that Weiss's (1970) specifications for recall are more or less encompassed within Lacks's (1984) system. For example, where Lacks (1984) would regard a design as sufficiently incorrect to award it a negative point, it will be rejected as insufficiently correct to be scored as adequately recalled. In addition, in developing a BGRT scoring system, focal issues surrounding the tenet of the gestalt, as already discussed, as well as established research literature will be incorporated. This will also render the two systems compatible across the three BGTs to be used in this research. The rationale and operationalization of a BGRT scoring system based on these principles will be presented in Chapter 7.

CHAPTER 5: ADMINISTRATION OF THE EXPANDED BGT

The method of administering the expanded BGT needs to be considered in detail for a number of reasons. Firstly, similar to other aspects of the BGT, there are a number of variations and no single standardized administration method exists. Field *et al.* (1982, p.838) note correctly that interpretation of research concerning the validity of the scoring systems on the BGT is problematic due, among other things, to the "nonstandard administration of the test". Others support this viewpoint (Billingslea, 1963; Dana *et al.*, 1983; Lezak, 1983; Olin & Reznikoff, 1957). While the administration of the BGT-*c* is fairly constant, issues such as the quantity of paper and the observation of subjects are important but seldom emphasized in the research literature. Secondly, particulars of administration of the BGRT-*i* lack detail. Thirdly, since only two studies could be found that used a BGRT-*d*, a rationale for administration procedures on this test needs to be developed and clearly outlined. Fourthly, the group method of testing has been decided upon (see 2.5: p.53), but Lacks's (1984) scoring system which was selected for this research (see 3.5.2: p.72) requires the observation of subjects. Therefore, certain adaptations are necessary to apply Lacks's (1984) methods to group testing. In addition, no literature could be found which applies the group administration method to any BGRT. Finally, the manner of administration needs to be carefully examined considering that the sample being tested consists of low-educated adults. Where research conducted on the BGT-*c* or BGRTs is lacking or insufficient, procedures followed in other neuropsychological tests will be considered. In summary, the administration of the expanded BGT will be carefully examined to reach decisions on the best manner in which to administer the BGT-*c*, the BGRT-*i* and the BGRT-*d* to groups of low-educated adults.

In what follows, administrative factors which are applicable to all three phases of the expanded BGT are discussed first. These are the changes in the actual nine BG designs, anxiety, eyesight and performance time. Then features specific to the BGT-*c* are reviewed. First the method of the presentation of the designs, with particular reference to group testing, is elucidated. Subsequently, the actual procedure involved in administration of the BGT is reviewed with emphasis on the importance of observation during the administration particularly with regard to rotations. Next, the administration of the BGRTs is presented, beginning with factors which influence recall. Next the procedures of administering the

BGRT-*i* and BGRT-*d* are reviewed. Finally, adaptations necessary for group testing will be discussed.

5.1 Factors pertaining to the administration of all three phases of the BGT

5.1.1 Variations of the BG designs

Although the BG designs are only presented on the BGT-*c*, they have obvious implications for the BGRTs. A large variety of the BG designs have emerged over the years. These range from slight to severe deviations from Wertheimer's (1923) original nine designs which Bender (1938) selected. Nearly four decades ago, Popplestone (1956) stated that this diversity of designs would provoke anxiety in psychologists who believed that precision is vital to science. However, there has been no progress regarding the use of a single set of designs and the assortment of designs continues.

Diversity exists in both (i) the number of BG designs used and (ii) in the designs themselves. Firstly, in considering the number of designs, the widely used Pascal-Suttell (1951) scoring system completely omits one design. An abbreviated form of the BGT-*c* utilizes six designs (Parsons & Weinberg, 1993), while Marsh (1972) replicated a study by Scorer (1964, in Marsh, 1972) because he had used only two of the designs which diminished the value of his results. Wagner and Marsico (1991) even suggested the use of a single design (specifically number 7) to indicate cerebral damage. Secondly, there are alterations of the angles, shapes and composition of the designs themselves. Popplestone (1956) notes that Bender (1938, 1946), who developed the test, is amongst those who have used designs which deviated from the original ones. For example, the BG designs on the cards she published in 1946 differed from those she originally presented (1938) and those in her manual of 1946 (Popplestone, 1956). Similar changes were made by other researchers such as Hutt, Wooltmann and Halpern, Pascal-Suttell and Peek and Quast, according to a review of design deviations by Popplestone (1956). Field *et al.* (1982) had to eliminate one scoring item as the design had been misprinted.

Billingslea (1963) stressed the need for a single, standard set of designs, adding that he had made this call for the first time in 1948. He made the supposition that clinicians' continued use of a variety of designs indicated that they believed the prerequisites of all the crucial tenets of perception were met, provided the fundamental gestalt of the designs were retained. Billingslea (1963, p.234) maintained such a belief was "untenable" and argued that if his supposition was correct, clinicians merited some of the "embarrassing criticism" they had received from research colleagues. Popplestone (1956, p.270) adds that the comparison of results of studies where such design disparity exists, becomes a "precarious procedure". Clearly, in order for the comparison of research findings on any version of the BGT to be meaningful, the designs used in the test administration should be limited to a single, standardized set. Further, the use of a single set of designs for any neuropsychological test is both logical and pragmatic. Lacks (1984), whose scoring system will be used, states that Bender's designs or Hutt's slightly modified ones may be used. Since this research aims at adhering to standardized procedures where possible, the designs originally selected by Bender (1938), which also appear in her manual (1946), will be used.

5.1.2 Anxiety

Chavez, Trautt, Brandon and Steyaert (1983) comment that although extensive research testifies that test anxiety influences test performance, this variable has been neglected by researchers. Chavez *et al.* (1983) found that performances were increasingly impaired as the level of trait anxiety reported by the subjects increased[1]. They caution that anxiety levels of subjects deserve consideration. Cotler and Palmer (1970) report that the performances of children who rated average on anxiety were as much affected as individuals who were rated as highly anxious. They postulate this result could be attributed to the fact that anxiety was compounded by the pressure demanded in the execution of the complicated procedures required of the subjects. They also note that anxiety slowed the speed of performance. Further, Finch Jnr, Anderson and Kendall (1976) found that high anxiety reduced the number of digits children were able to recall on the Digit Span Test. They attributed this to impaired concentration caused by high anxiety. It is interesting to note that Tolor (1956) found scores on the BGT to be less susceptible to anxiety than those of the Digit-Span Test. Bearing the above in mind (see 6.6: p.111) anxiety could be a confounding variable in a low-educated sample group. For this reason, a ninth observational item of extreme anxiety will be added to Lacks's (1984) nine observational criteria (see

---

1. These levels were measured by the A-Trait from the State-Trait Anxiety Inventory.

Appendix 3: p.257). Thus, where extreme anxiety is evident, the test scores of subjects will be excluded from the normative data collected in this research.

5.1.3 Eyesight

Kempen, Kritchevsky and Feldman (1994) highlight that impaired eyesight can jeopardize neuropsychological test results. Subjects who fall into a low-education group, such as those in this study, have an increased likelihood of impaired, uncorrected eyesight. The reason for this is that low-educated subjects will most likely fall into a lower socioeconomic and sociocultural group (the reasons for this supposition will be discussed in more detail later (see 6.6: p.111). In such individuals, uncorrected, impaired eyesight, could be attributable to either lack of awareness of poor eyesight or the absence of financial means to correct such a problem or both. An example of possible confounding is evident in McCarthy's (1975) study. He projected the BG designs onto a screen to groups of up to 30 subjects, without any prior evaluation of the subjects' eyesight. Because of this method, errors in copying the designs might have been due to an inability to see the designs properly, rather than perceptual-motor deficits. Within the scope of this research it is not possible to conduct an eye test on every individual, yet it is clear that this factor is an important one and merits consideration. However, certain measures can be taken to implement a degree of control over this variable. Firstly, indications of impaired vision will be added to Lacks's (1984) list of observational criteria (see Appendix 3: p.257). The scores of subjects who indicate verbally that they are having trouble seeing the designs, or those whose behaviour indicates this, will be excluded from any normative data. Secondly, this factor will be strongly considered when choosing the procedure for the presentation of the BG designs (see 5.2.1: p.96).

5.1.4 Time taken

Lezak (1983) notes that patients with right hemisphere damage are inclined to perform neuropsychological tests more slowly. Ponsford and Kinsella (1992) add that patients suffering from neurological damage are inclined to forego speed in favour of preserving accuracy. Generally, no time limit is imposed on subjects doing the BGT-*c* (e.g. Koppitz, 1975; Lacks, 1984; Marsh, 1972; Thomas, 1984; Yousefi *et al.*, 1992). Nevertheless, the time taken to complete the BGT-*c* has been thoroughly researched, with particular regard to investigating the presence of cerebral damage (Andert *et al.*, 1978; Lacks, 1984; Rosecrans & Schaffer, 1969). In some cases, time has emerged as a significant factor in the differentiation between organic and emotional pathology (Andert *et al.*, 1978; Lacks, 1984). Lacks (1984, p.25) has an optional scorable error in her scoring system, for a subject taking more than 15 minutes to complete their protocol as this is "usually another strong organic sign". Similarly, Andert *et al.* (1978) found that the longer the time a subject took to copy the designs, the poorer the protocol was. In contrast, Rosecrans and Schaffer (1969) found that time as an indicator of cerebral damage, was inconclusive.

While Lacks (1984) regarded time as significant enough to include as an optional scoring error (as mentioned above), she used the individual testing method. However, in this study the possibility of group dynamics and the education level of the subjects must be considered. Regarding group dynamics, this author observed that during prior research with children (Dyall, 1993), this factor played a role in the performance time. Most individuals in a group finished within a few minutes of each other regardless of the time taken, indicating that subjects kept pace with each other albeit consciously or unconsciously. This indicated the presence of a possible competitiveness or concordance between the subjects. Alternatively, this might be attributable to a group which is habituated to doing written work together, such as copying information from a blackboard. Supporting this was the fact that the children were tested according to their school classes and therefore worked with each other daily. These group dynamics might or might not be relevant to groups in general. Only further testing of different groups, such as those in the current study, will elucidate the real reasons for this effect. A further, essential consideration concerning the time factor in this particular study, is that subjects with a very low education, or without any education at all, are likely to be unfamiliar with and unaccustomed to even holding a pencil or using an eraser. Because of this, the time taken by such subjects to actually copy the BG designs is likely to be longer than it is for more educated subjects. Also, as already discussed (see 5.1.2: p.93), anxiety is likely to be present in some subjects and research has shown this can slow performance speed (Cotler & Palmer, 1970).

Similar to the BGT-*c*, most researchers generally do not mention imposing a time limit on the BGRT (Armentrout, 1976; Hutton, 1966; Imm *et al.*, 1991; Peek & Olson, 1955; Rogers & Swenson, 1975; Tolor, 1956, 1958; Weiss, 1970). In contrast, Goodstein *et al.* (1955)

imposed an eight minute time limit. However, research has found that additional time on the BGRT does not improve performance (Edwards, 1980).

Regarding the method of timing subjects, Andert *et al.* (1978) used a stopwatch. However, Lacks (1984) advises against this as pinpoint accuracy is not necessary and a stopwatch may increase anxiety and be intrusive to testing. Koppitz (1975) also noted that subjects under time pressure can become anxious.

In light of the above discussion it was decided that no time limit would be imposed for any of the three tests. Further, because of a likely but unknown relationship between the education of subjects and the time they take to complete the BGT-*c*, it was decided that Lacks's (1984) optional scoring error would not be applied to the current research. It was nevertheless decided that the time taken by each subject to complete their protocols would be recorded. The reason for this is to collect information on the mean performance time for low-educated subjects who are tested according to the group method. It is envisaged that reporting these times could later assist assessment by providing an additional diagnostic clue, along with observational cues, in borderline or problematic cases. Once again, considering the available literature, care will be taken to time subjects unobtrusively.

5.2 Administration method: the BGT-*c*

5.2.1 Variations in the presentation of BG designs with particular reference to group testing

The individual presentation of the BG designs requires that the examiner produces each of the nine cards on which the designs are printed, one at a time, to the subject (Koppitz, 1975; Lacks, 1984). However, this method is not practical for group testing. Different methods of group administration on the BGT-*c* (Koppitz, 1975) include the presentation of a single set of enlarged cards, the use of individual copy books each containing the set of designs, projection of the designs onto a screen, and a set of cards for each subject. The method must be carefully chosen in order to prevent confounding variables from influencing the test results. As already discussed (see 5.1.3: p.94), low-educated individuals from a lower socioeconomic level are more likely to have impaired eyesight than a more educated, financially prosperous population. Research was reviewed which showed that poor eyesight can confound test scores. It was therefore concluded that the administration procedures should consider this possibility carefully.

Thus, for this research, the overhead projector method and the presentation technique of using a single set of enlarged cards were rejected. The copy-book procedure was rejected as it is not cost effective and the aim of this study is to provide additional information in a manner which is economical in terms of both financial cost and time. The individual card method is closest to the original manner of design presentation and appears to be the most appropriate for group testing. For the reasons argued above, it was decided to use this method of design presentation.

5.2.2 Procedure

A quiet place is needed for testing, where there will be no disturbances or distractions. In this regard, Patton and Offenbach (1978) found that children made more errors on visual and auditory memory tasks when visual or auditory distracters were present.

Lacks (1984) advises the tester to ensure that the writing surface is smooth, presumably to avoid an uneven surface from mimicking motor incoordination errors. It is essential for the cards to be correctly ordered and orientated (Lacks, 1984). Regarding the *quantity* of paper, Lacks (1984) and Koppitz (1975) both agree that the quantity of paper used should not be restricted. However, some researchers allow only a single page of paper (Dibner & Korn, 1969; Smith & Martin, 1967; Yulis, 1970) while others require subjects to draw a single design per page (Keogh & Smith, 1961). This latter method of administration eliminates valuable information regarding the spatial orientation of the designs and their relation to each other. Dibner & Korn (1969) admitted that one of Koppitz's emotional indicators could not be scored because subjects were restricted to a single sheet of paper. Also, the close proximity of designs is a scorable error on the Lacks (1984) system, which will be used to evaluate the BGT-*c*. The *size* of the paper used in the administration of the BGT-*c* is also important. The standardized paper prescribed by Bender (1938) is blank A4 paper. Research shows that alterations in paper size can alter test performance. For example, Allen and Frank (1963) found fewer errors on BGT-*c* protocols when the paper size was the same as the that of the cards bearing the designs. In light of the above, it was decided that subjects in this research will be allowed to use as much paper as they wish and the standard A4 paper size will be used.

Lacks's (1984) administration involves aligning the paper in a position vertical to the subject. However the subjects may reorientate it to suit themselves. The BG designs are placed parallel with the top of the paper. Subjects are asked to draw exactly what they see. Mechanical aids such as rulers or coins are not allowed but the use of erasers is permitted. Questions are countered by repeating the instructions. The examiner reorientates a card if the subject rotates it. A subject who persistently rotates the card is left to do so, but the protocol is marked at the end of the test to indicate the top of the design as it was drawn. Observational cues, according to Lacks's (1984) observational criteria (see Appendix 3: p.257) are made and the time taken to complete the protocol is noted. Various researchers have made additions to the administration procedure, including the introduction of practice effects and learning cues to correct rotations (Smith & Martin, 1967). Yet no addition is quite so unacceptable as that of Yulis' (1970), who included the threat of a 90 volt shock, to be administered during the execution of the BGT-*c*. It was merely a threat to one sample group, but was actually administered to subjects in another. He abandoned this method later in his research as he experienced difficulty procuring subjects! In order to adhere to a standardized procedure, no additions to the administration procedures will be made, apart from those necessary for group testing (see 5.4: p.103).

5.2.3 Observation of subjects

Observation of behaviour is critical to ensure correct evaluation and diagnosis of the BGT protocols (Koppitz, 1975; Lacks, 1984). Koppitz (1975) notes that through observation alone, much information can be gained and important differentiations can be made. Characteristics such as aggression, impulsivity, compulsiveness and anxiety can influence performance and these should be noted for effective interpretation of the BGT-*c* (Koppitz, 1975). Lacks (1984) documents that a subject who hardly glances at the designs and completes the protocol in two minutes, is unlikely to have cerebral impairment, regardless of the score. Conversely, excessive effort and deliberation over a long time period could result in a protocol which is just below the organic 'cut-off' point. She warns that the administrator must try to establish whether errors have occurred as a result of confounding factors such as those listed in her observational criteria (see Appendix 3: p.257) or because of genuine perceptual difficulties.

Lacks (1984, p.23) warns that some examiners are "lulled into complacency by a deceptively simple test to administer" and do not observe closely enough.

Despite the importance of observations, many studies have simply used protocols obtained from existing records (Cooper *et al.*, 1967; Dierks & Cushna, 1969; Field, *et al.*, 1982; Friedt & Gouvier, 1989; Griffith & Taylor, 1960; Jernigan, 1967; Johnson *et al.*, 1971; Marsico & Wagner, 1990; Morsbach *et al.*, 1975; Parsons & Weinberg, 1993; Wagner & Flamos, 1988) without any mention of observations or observational notes. This is presumably because these do not exist. Field *et al.* (1982) actually reported being unable to score two items as the administration had not been observed, while Freed (1966) stated that it was difficult to assess certain rotations as the examiner's notes were unclear as to whether or not the rotational errors on the protocols were due to rotation of the cards by the subjects. The lack of adequate observation in these two studies is a serious omission considering that both studies were investigating the significance of rotations. This highlights another important reason for careful observation - namely to establish the reason for rotations.

5.2.4 Rotations

The significance of BG rotations, as an indicator of neurological impairment, has already been detailed (see 4.2.5: p.81) and rotations are also regarded as a diagnostic sign (see 2.1.5: p.12). In view of this, it is of critical importance to observe whether rotations are executed due to a technical or a perceptual error. For example, a subject may turn the paper or card upside-down but copy the design correctly from this orientation, or he or she may rotate the design in the process of copying it from the card because of perceptual-motor distortion. Yet only a few researchers note that care should be taken regarding the orientation of the paper, namely parallel to the designs (Koppitz, 1975; Lacks, 1984; Smith & Martin, 1967). Sabatino & Ysseldyke (1972) held both the cards and the paper to prevent rotations from occurring, while Snyder and Kalil (1968) ensured that the card was not turned but subjects were permitted to rotate the paper.

Lacks (1984) identified rotations performed by subjects due to the cards being turned, by marking the protocol at the end of the test. However she tested subjects individually. In group testing it is problematic for the tester to remember each instance of rotation until the end of the test administration. This method is therefore not feasible for group testing. In previous research (Dyall, 1993), subjects were warned that the testers might draw a line while they were working (this was demonstrated before testing began), but that this should not concern them. This method proved to be effective. While this might interfere marginally with test performance, it is the most practical option. It is far less disadvantageous than scoring a rotational error, when in fact, the card has been rotated and the design correctly copied. Considering the importance of rotations, it was decided to adhere to this method in the current research.

5.3 Administration methods: the BGRTs

Available literature on the BGRT-*i* is considerably less than that on the BGT-*c*. This is in keeping with the statistics of Schulberg and Tolor (1961) which reveal this version is less popular than the BGT-*c*, as already discussed (see 2.4.2: p.48). However, administration procedures appear to be simple and very similar in the studies which have been done. Regarding a BGRT-*d*, only two references could be found (Edwards, 1980; Goodstein *et al.*, 1955) which contain no BGRT-*d* administrative procedures. It was thus decided to source literature concerning the administration of the BGRTs from other memory tests and specifically information pertaining to procedures followed in other tests of visual memory.

5.3.1 Factors influencing recall

Since there is no single method of administering the BGRTs in the studies reviewed, two decisions need to be made - namely, whether the recall should be incidental or intentional and whether the tester should offer encouragement to the subjects.

(i) Spontaneous versus intentional recall

Spontaneous or incidental recall tasks involve requesting the recall of designs without prior warning (Lezak, 1983). Concerning spontaneous versus intentional recall, no significant difference between incidental and intentional recall were found on a copy and recall test - the Key Osterreith Complex Figure (Dinklage, Kantor & Grodzinsky, 1992). Some studies on the BGRT-*i* have used the incidental recall method (Peek & Olson, 1955; Rogers & Swenson, 1975), while numerous others make no mention of which recall procedure was used (e.g. Armentrout, 1976; Holland & Wadsworth, 1979; Lyle & Gottesman, 1977;

Sabatino & Ysseldyke, 1972; Tolor, 1958). Anticipation of the BGRTs might distract the attention and/or concentration of the subjects from the BGT-*c*. In particular, with the subjects in this study, an intentional recall test might also create anxiety, which has already been discussed as a possible confounding variable (see 5.1.2: p.93). In this regard, Cooper and Sagar (1993) found that Parkinson's disease victims made more recall errors with intentional recall than with incidental recall. They attributed this difference to the stress that is implied in intentional recall. For the reasons given, an intentional recall could exert a confounding influence on the scores. This research aims at adhering to standardized procedures, where these exist. Thus it was decided, for the purposes of this study, that the subjects will not be prewarned of the BGRTs.

(ii) Urging subjects to recall designs

Hutton (1966) repeatedly encouraged his subjects to attempt to remember further designs. Similarly, Rogers and Swenson (1975), urged children who indicated they could not recall any more designs, to take more time and attempt to remember more. Tolor (1956) asked subjects who had completed the recall to draw any part of a design which they could recall. Armentrout (1976) used an even more persuasive method. Encouragement was followed by the suggestion that parts of a design be drawn, and after that, subjects were even asked to make guesses with their recall attempts. However, Edwards (1980) found that positive feedback to subjects created a willingness by subjects to draw designs which they could not recall well enough to draw accurately, thereby increasing errors. In a group context, encouragement is difficult to implement as each individual works at his or her own pace and subjects are thus in various stages of completion of the three tests. Further, encouragement is a difficult component to replicate precisely in other studies. Without such standardization, an instructional set involving encouragement could become a confounding factor in test results. Previous research by the present author (Dyall, 1993) found that subjects were quite adamant as to their recall ability, without any persuasion being offered. This is illustrated by the fact that subjects were equally resolute about stating that they could not remember any more configurations or that they required more time to remember more designs. It was therefore decided not to encourage subjects.

5.3.2 Procedure: the BGRT-*i*

In immediate recall tasks on other neuropsychological tests, subjects are simply asked to remember what they can after the initial exposure to the designs (Lezak, 1983). In the limited literature available on the administration of the BGRT-*i*, some researchers did not mention the administration procedures they used (Clark, 1982; Collaer & Evans, 1982). Those who did simply asked subjects to remember and draw as many designs as they could once they had completed the BGT-*c* (Finch Jnr, *et al.*, 1983; Hutton, 1966; Imm *et al.*, 1991; Peek & Olson, 1955; Rogers & Swenson, 1975; Tolor, 1956; 1958; Weiss, 1970). However, Tolor (1958) added the instruction that subjects should draw as many *parts* of the designs as they could remember. This is contrary to the basic tenets of the BGT as it encourages the fragmentation of the gestalt. Thus, in keeping with the majority of the literature available, on the completion of the BGT-*c*, subjects will be asked to draw as many of the designs as they can remember. As the incidental manner of recall has already been decided upon above (see 5.3.1:i p.100), no prior warning of the recall test will be given. Adhering to a prior decision made above (see 5.3.1(ii): p.101), no encouragement to recall additional designs will be given. In order to allay anxiety, discussed (see 5.1.2: p.93), which is often evident at this stage and can be considerable, questions such as "How many designs were there?" would be answered by saying "You might not be able to remember them all. Just draw the ones you can remember."

5.3.3 Procedure: the BGRT-*d*

As already stated, only two studies on the BGRT-*d* could be found (Edwards, 1980; Goodstein *et al.*, 1955). Both studies administered a BGRT-*d* after an initial exposure to the designs, which varied in form between the two studies, and once the BGRT-*i* had been executed (Edwards, 1980). In a review of delayed recall tests, Lezak (1983) notes that the immediate and delayed tests are interspersed with an interference activity which takes the form of another test not involving memory. She reports that the delay can vary from five up to ten minutes. With the Complex Figure Test, the delay between the immediate and delayed memory phases varies between 20 and 45 minutes (Lezak, 1983). There appears to be little significant difference between time delays which are less than an hour in duration (Lezak, 1983). Goodstein *et al.* (1955) first administered the tachistoscopic version of the BGT and then a BGRT-*d* after six minutes. The tachistoscopic method used by Goodstein *et al.* (1955) was rejected for the present research in favour of the BGT-*c* (see 2.1.6.4: p.19). The six minute interval used by Goodstein *et al.* (1955) is a rather short for the completion of the Draw-A-Person test. In addition, 10 minutes is a concise time period and facilitates timing, which exercises a precise degree of concentration on both the researcher and/or the research assistant, when testing groups of people.

It was therefore decided to conduct a BGRT-*d* which would begin 10 minutes after the completion of the BGRT-*i*. It was also decided that the instructions used on the BGRT-*i* would be repeated and, again, neither prior warning of this task nor encouragement would be given.

5.3.4 The Goodenough Draw-A-Man Test as an interference task

Bender (1946) often interspersed her testing with the Goodenough-Draw-a-Man Test, and other researchers have followed her example (e.g. Goodstein *et al.*, 1955). This test was introduced in 1926 and remained unchanged until Harris modified it in 1963 (Anastasi, 1976). The Goodenough scale has been widely used in the clinical assessment of intellectual maturity in children and many clinicians have calculated a mental age from it (Harris, 1963). In children, an estimate of IQ is obtained by dividing this mental age by the chronological age and multiplying by 100, and is an indication of a child's rate of development (Cronbach, 1991). The test was developed by analysing developmental changes occurring in over 900 spontaneous drawings from children. Harris (1963) modified the scoring procedure, but Vane (1967) states this is questionable, so Goodenough's original version (Anastasi, 1978) will be adhered to. Since this is being used as an interference test only, these scores will not be reported.

5.4 Administrative considerations relevant to group testing on the expanded BGT

Certain adaptations need to be made to the group testing method in light of the administration decisions which have been made above. Firstly, the importance of observation, particularly when using the Lacks (1984) scoring method has already been discussed (see 5.2.3: p.98). For the purposes of the present study therefore, it was decided to use the group method of testing, with a limit to the number of subjects per group. This limit ensures that the group is small enough to obviate relinquishing the clinical significance of observation during the test administration. Drawing on this author's previous experience of group testing, it was decided that a maximum of 14 subjects per group would be tested with the researcher and a research assistant present, and 8 subjects per group when the researcher worked alone.

It was decided that the individual card method would be used in this study (see 5.2.1: p.96). However, since this method is being applied to the group administration, instead of the tester presenting the cards, the subjects will be given their cards and asked to begin with the first design and work through the pack, turning their cards as they complete each one. From this author's experience this necessitates allowing for time to check the cards after each administration, especially since the individuals turn each card over themselves. The cards can often get mixed up or replaced upside-down. It was also decided (see 5.2.2: p.97) that subjects would be allowed to use as much paper as they wish. For group testing, in order to facilitate free access to paper, subjects will be given several pages each, which will be replenished as necessary. As discussed (see 5.2.4: p.99), it is considered less disadvantageous to a subject to note rotations made due to the incorrect orientation of a card or the paper, rather than to score this as an error. Thus, when an individual copies a design from a rotated card, the author or research assistant will mark the baseline angle of the card on the protocol. Subjects will be warned prior to the onset of testing that the researcher(s) might mark their paper while they are working, without being told why this might be done.

During the BGRT's using the group testing method, it is necessary to space subjects sufficiently far from one another so that designs drawn by one subject are not visible to another. This is particularly important as one subject may be working on the recall phase and another may still be working with the original set of designs. Where space is limited, screens between subjects will be used instead. Also, because subjects have a pile of paper, it is necessary for the researcher(s) to ensure that where design imprints appear on sheets of paper below, these are removed before a BGRT begins.

The rationale of the scoring systems and method of administration have been discussed and decisions have been made as to the procedures which will be applied in this study. The only remaining consideration is the selection of subjects which is now addressed.

CHAPTER 6: SELECTION OF SUBJECTS

In the selection of subjects for this research it is relevant to consider factors which influence performance on neuropsychological tests and where necessary to account for these in the research design. The following discusses each variable under its own heading in the following sequence. First, research findings from studies on neuropsychological tests in general, for the variable in question, are discussed. This is followed by findings concerning the relevant variable's applicability to the tests which make up the expanded BGT. The discussion of each variable concludes with a decision regarding the way in which each particular factor is accounted for in this research.

6.1 The importance of subject variables

The importance of subject selection and accounting for specific variables is highlighted by research on both neuropsychological tests in general and on the BGT in particular. Anthony, Heaton and Lehman (1980) found that demographic variables significantly influenced performance on the Halstead-Reitan Battery. On the BGT-*c*, Adams *et al.* (1982) concluded from research that confounding variables exerted such an influence that those with certain personal characteristics would be classified as brain damaged regardless of their neurological status. Undoubtedly variables which influence performance of the expanded BGT are of critical importance.

6.2 Gender

Some neurological tests reveal gender differences. For example, the Finger Tapping Test (part of the Halstead-Reitan Battery), revealed that normal males scored higher than their female counterparts, whose mean score fell into the neurologically impaired range (Chavez *et al.* 1983). Also, Dodrill (1979) found notable gender differences on motor and/or visuospatial tasks, on the Halstead-Reitan Battery and on other neuropsychological tests. Leckliter and Matarazzo's (1989) review of the same battery confirmed certain gender differences. On psychomotor tasks, Houx and Jolles (1993) found that gender differences were less significant between neurologically impaired individuals than between normal subjects. The authors concluded that gender differences decrease in significance as neurological impairment becomes more severe. Lacks (1984) supports this view.

Conversely, on the BGT-*c*, a gender effect does not appear to be clinically significant. In young children, Koppitz (1962, 1975) cites that her own and others' research reveals that gender does not influence performance on the BGT-*c*. Studies by Fiedler and Schmidt (1969) and Gilbert and Levee (1967) confirm this. These findings also hold for retarded children (Condell, 1963; Doubros & Mascarenhas, 1969). The lack of gender differences remains consistent across ethnic groups. Taylor and Thweatt (1972) found no evident differences between the scores of boys and girls in both Navajo Indian and Anglo subject groups. Welcher *et al.*'s study (1974) of ethnically mixed children support these results.

Within a South African sample of Zulu-speaking children, Viljoen *et al.* (1994) found that gender and geographical factors combined accounted for only 3% of the total variance of scores across a sample of 882 subjects. However, the authors emphasize that the severe lack of resources in rural schools together with other possible confounding variables could have affected these results. Viljoen *et al.* (1994, p.149) comment that the deficiency of "resources" at all the schools in which they conducted testing was particularly prevalent at rural schools. For these researchers, this lack confirmed the sub-standard level of education in rural schools. Thus the combined 3% variance for gender and geographical area could be due to the low educational standard in rural schools rather than to any gender effects. If this factor could have been controlled, the variance might have been lower than 3%.

In adults, Lacks (1983, p.60) found that gender in non-patient adults produced a "small but statistically significant" effect, using her own scoring system. She adds, however, that other studies using the same system did not. For example, Brilliant and Gynther (1963, p.476) found that gender differences on the BGT-*c* in a psychiatric inpatient sample of adults were "negligible", as did Miller and Hutt (1975). Lacks (1984, p.46) attributed these differences to her large sample numbers and stated that gender differences overall had "no clinical significance". She did not separate her data according to gender. Studies using other scoring systems in mixed inpatient adults supported the absence of gender effects (Carlson, 1966; Pascal & Suttell, 1951).

Contrary to other literature reviewed, a single study indicated gender differences on the BGT-*c* (Dierks & Cushna, 1969). However, this study was somewhat confusing, stating initially that no significant gender differences were revealed. Yet Dierks and Cushna (1969) later concluded that gender was a major influencing factor. It appears that the authors compared performance according to each individual Koppitz error criterion and found gender differences in the individual *types* of errors made. Further, the reliability of the study was questionable as it used the Koppitz Developmental Scoring System beyond the applicable age range. While these contradictory findings are noted, it must been seen within the context of a single study which contrasts to many on the subject.

Regarding memory, Dodrill (1979) found that across a number of neuropsychological tests, no significant gender effects were revealed which involved memory, apart from on the Wechsler Memory Scale, Logical Memory. Snow and Sheese (1985) failed to find gender differences in the memory abilities of brain damaged individuals. Regarding memory for designs, which as already discussed (see 2.4.1: p.46), is pertinent to the BGRT, Dinklage *et al.* (1992) found no significant gender differences on the recall of the Complex Rey Figure. Further, Brilliant and Gynther (1963) established that two memory tests, the Graham-Kendall Memory-for-Designs and the Benton Visual Retention Test revealed no significant differences between genders. On the BGRT, Edwards (1980) found no gender differences in children.

Lacks (1984) did not stratify her normative data for gender and her scoring system is being used in this research. Concluding from the above, it was decided that in this study no attempt will be made to control for gender differences.

6.3 Intelligence quotient

The relationship of BGT-*c* performances to IQ appears uncertain. Cerbus and Oziel (1971) found no significant relationship between IQ and BGT-*c* performances in children. Similarly, Koppitz (1975) reported that good performance on the BGT-*c* of both culturally disadvantaged children and non-white learning disabled children correlated well with the rates of advancement, but not with their IQ scores (Koppitz, 1975). Koppitz (1975) concluded that the test is more related to performance than IQ. With Lacks's system (1984) correlations of r=-0,33 and r=-0,13 were found between IQ scores and the BGT-*c* (Brilliant & Gynther, 1963; Lacks, 1984). Lacks's (1984) scoring system was used to evaluate the BGT-*c*, and the system operationalized in this thesis was used to score the BGRTs. Other studies have also found a correlation between IQ and BGT-*c* performance (Pascal & Suttell, 1951). Considering race and IQ scores, non-white learning disabled pupils obtained good scores on the BGT while scoring comparatively low on verbal IQ tests (Koppitz, 1975). Dyall (1993) reported a correlation between performance on the *expanded* BGT and IQ in children of r=-0,24. Lacks's (1984) scoring system was used to evaluate the BGT-*c* and the system operationalized in this thesis was used to score the BGRTs. Regarding the BGRT-*i*, Aaronson (1957, p.187) concluded from research that no "practical relationship" between IQ and this test existed.

While it is impossible within the scope of this study to administer IQ tests, the literature suggests that the influence of IQ on performance is relatively low. In addition, some degree of a normal IQ distribution will be obtained by testing people from varying occupations such as electricians, farm workers, security guards, restaurant workers and retail employees. Thus the influence of IQ, if any, will be reduced.

6.4 Age

Leckliter and Matarazzo (1989) found age effects on the Halstead-Reitan Battery, with younger individuals performing better than older subjects. On the BGT-*c*, Adams *et al.* (1982) warn that age may confound test scores. Lacks (1984) endorses this and recounts that the interrelationship between perceptual-motor ability and age is broadly recognized. Age has a significant effect on the BGT-*c* protocols of children (Keogh & Smith, 1968; Koppitz, 1975; Taylor & Thweatt, 1972; Viljoen *et al.*, 1994; Yousefi *et al.*, 1992). For this reason, Koppitz (1975) provides clearly age-stratified normative data for children on her Developmental Scoring System. However, it must be noted that the strong age effect on the Koppitz (1975) scoring system is affected by two other variables, namely, developmental and educational levels. Firstly, Koppitz's (1975) system is specifically designed to establish developmental levels, which change rapidly in children as age increases. However, since Koppitz (1975) states that developmental levels stabilize around 10 years 11 months this factor is unlikely to have as strong an influence on the scores of adults. Further, other scoring systems, such as that of Lacks (1984), are designed to detect neurological impairment instead of developmental factors. Secondly, it must also be borne in mind that education increases with age in children, while this is not normally true for mature adults. Because of the link between age and education in children, studies with children can be expected to reveal stronger age effects than those with adults.

In a South African population of Zulu-speaking school children, Viljoen *et al.* (1994) found that age accounted for 31% of the variance in scores. They concluded that age is "clearly the most important variable" (1994, p.149). However, a serious omission in this study is that education effects were not considered. This author postulates, in light of the discussion above on the relationship between age and education, that a high percentage of the variance attributed to age by Viljoen *et al.* (1994) could in fact be attributable to education. In adult samples, Tolor and Brannigan (1980) infer that age in non-patients is only significant prior to maturation and again, once cerebral decline attributable to aging commences. This concurs with other research. Pascal and Suttell (1951) restricted their age range to 15-50 years because no significant age effect on the BGT-*c* existed within this range. Lacks (1984) found an age effect, particularly in older adults. In her sample of non-patients, the lowest age group of 17-24 years had a mean error score of -1,47, which increased to -2,62 for the 55-59 year old age group. In a study of older adults, the mean score for the 60-64 year old age group was -3,23 which increased to -4,33 for subjects over 80 years of age (Lacks & Storandt, 1982). Only 7% of non-patient subjects aged between 45 and 59 years scored within the organic 'cut-off' range. This increased sharply to 22% for the 60-74 year old age group and to 39% for adults over 80 years of age. In all, Lacks (1984) reports a correlation of 0,30 ($p < 0,01$) between age and performance on the BGT-*c*. She concludes that the scores of non-patient adults aged from 17 to 59, with no prior psychiatric or neurological impairment, are "consistently" within the non-organic range (Lacks, 1984, p.46).

Considering memory, Hutton (1966) found that improved BGRT scores in children were significantly related to age. This is in keeping with the discussion of the correlation between age and education effects in children as opposed to adults. This supposition is further confirmed by Reznikoff and Olin (1957) who established that age did not have a significant effect on the recall of BG designs within a 18-49 age group.

Thus, it appears that particularly significant differences occur in the adults in the age group above 59. Considering this, adults older than 59 years will be separated from those who are younger.

6.5 Race and ethnocultural effects

In South Africa, education has been compulsory for whites for many decades and, until recently, was available at a minimal cost or free for this racial group. Consequently all educable white individuals received an education. In contrast, for the 'other-than-white' group, education has been neither compulsory nor always economically or geographically accessible. For this reason, the subjects in this study (non-clinical adults with less than 9 years of education), perforce, constitute 'other-than-white' individuals. Additionally, as Xhosa is the dominant cultural group in the Eastern Cape, the subjects will be predominantly Xhosa-speaking.

From the above, it follows that the choice of the educational level in this study inherently predetermines race and ethnocultural factors and therefore controlling for these factors as influencing variables is not possible. However, race and ethnocultural factors become pertinent when the preliminary normative data presented in this research is applied to other racial or ethnocultural populations with similar education levels. The issue will therefore be discussed briefly.

Race and cultural effects can influence performance on neuropsychological tests (Bornstein, 1986). In some instances, racial and/or ethnic factors have also been found to influence performance on the BGT-*c*, particularly when certain scoring systems are used. For example, significant differences between races were found when using the Pascal-Suttell (1951) scoring system or a modified version thereof (Adams *et al.*, 1982; Butler *et al.*, 1976; Carlson, 1966). Using the Koppitz system, Yousefi *et al.* (1992) found that Iranian children scored lower on the test than their American counterparts, and other studies which also used this method found differences (Cerbus & Oziel, 1971; Henderson *et al.*, 1969; Isaac, 1973). However, factors other than racial or ethnocultural differences could also be involved. Carlson (1966) states IQ might have confounded racial differences and Butler *et al.* (1976) also suggest from their own research, that racial differences found in previous studies could have been due to IQ instead.

The Lacks (1984) system does not appear to have a strong racial bias. No racial differences were found using the Hutt-Briskin system (Brilliant & Gynther, 1963). As with gender, Lacks (1984) found small degrees of statistically significant differences between race in large samples, when using her own system. However, she stated that these differences were not of clinical importance (Lacks, 1984). She therefore did not separate her normative data according to racial groups. Research which used Lacks's (1984) BGT-*c* scoring method and the BGRT system presented in this research to evaluate the expanded BGT protocols of racially mixed children, found that no significant racial differences existed (Dyall, 1993). From the above literature, it appears that the normative data obtained on the expanded BGT, using the above scoring systems, could possibly be applicable to other racial and cultural groups.

6.6 Socioeconomic and sociocultural effects

It can be logically deduced that individuals with no or little education constitute a lower-than-average sociocultural and socioeconomic level because they are unlikely to have sufficient vocational skills to alter their economic situation. Frey and Pinelli Jnr (1991) found that the socioeconomic level of school children influenced the BGT-*c* scores even more than their school grades did. However, Isaac (1973) found no significant differences between the performance of advantaged as opposed to disadvantaged children. Yousefi *et al.* (1992) investigated the scores of Iranian children using the BGT-*c* which was scored according to the Koppitz Developmental Scoring System. He describes Iran as "a non-Western developing country" (Yousefi *et al.*, 1992, p.411). Means in this study are significantly lower than for Koppitz's normative data obtained from samples of American children. Yousefi *et al.* (1992) propose some interesting possible influencing factors to explain their results. Firstly, the Iranian culture has traditionally ignored areas such as the graphic arts and the teaching of visuomotor skills in Iranian schools. This trend has been exacerbated since the Islamic Revolution when art lessons were restricted to instruction in Arabic-Persian calligraphy. Zuelzer, Stedman and Adams (1976, p.875) also found significant differences in different socioeconomic groups and they concluded that these differences indicated the need for "specific" normative data. Welcher *et al.* (1974) studied the BGT-*c* scores of inner-city American children from lower-middle and lower socioeconomic families. These were significantly lower than the Koppitz norms and the researchers maintained that the children's visuomotor development was retarded for their age. However, they concluded that this retardation might be attributable to factors other than neurological damage such as environmental conditions which included little intellectual stimulation, or even the unfamiliar testing situation.

In conclusion, lower socioeconomic and cultural levels may possibly negatively affect the performances of individuals from these backgrounds on the BGT. The research reviewed suggests that a number of influencing factors could be involved. These include the tradition extant in particular cultures, which in turn affects not only the *way* in which schooling is conducted, but its *emphasis* as well. Further, the degree of stimulation and familiarity with the test situation is pertinent. It must be acknowledged that low-educated adults, such as those in this study, could be affected by a number of the above conditions. Firstly, their childhood environment is more than likely to have been disadvantaged. Secondly, the schooling system, which was operative until very recently, has been stringently criticized for its bias in favour of traditionally 'white' schools and inadequacy of education in traditionally 'other-than-white' schools. Thirdly, the lack of mental stimulation possible during their childhood, and later occupations in mostly unskilled or semiskilled jobs, is also a factor to be considered. Finally, a point which must be unequivocally emphasized is that a low-educated person is unlikely to be familiar with holding a pencil, writing, sitting at a desk, or with the comparatively formal conditions of the testing environment and the visuomotor experience of drawing and copying.

It was decided to apply a dual approach to consider the factors discussed. Firstly, it was decided that the tester should make an effort to help subjects to feel as unthreatened as possible, and where at all feasible, to test subjects in surroundings familiar to them. Secondly, it was anticipated, that the expanded BGT scores for this sample of adults could possibly be lower than those for relatively advantaged children with the same levels of education. Children undergoing schooling are familiar with the classroom situation and with writing, and are exposed to stimulation at least during school hours. Anxiety as a confounding variable is closely related to the above and has already been considered (see 5.1.2: p.93).

6.7 Neurological impairment, medication and substance abuse

Lacks (1984) states that her normative data is applicable to individuals without a history of psychiatric or neurological impairment. Neurological impairment can be caused by a variety of factors which included diseases with cerebral involvement and traumatic head injury (see 2.3.4: p.40). Sedative medication can decrease the concentration and alertness of individuals. Further, Miller and Branconnier (1983, p.453) state from their own and other research findings that memory dysfunction is "the single most consistently reported psychological deficit produced by canabinoids". The nature of memory impairment due to cannabis is comparable to that in individuals with neurological dysfunctions and appears to be related to impairment in the cholinergic limbic system. Three reefers a day is considered to be an intake at which such side-effects become significant (A. Levin, personal correspondence, May 1994).

Considering the above, any person with a history of unconsciousness, an open head injury or a disease which is neurologically implicated such as meningitis, encephalitis, and epilepsy will be excluded from the sample. It was also decided that the protocols of subjects who have ever taken psychiatric medicine (i.e. subjects with a psychiatric history) and those currently taking sedative medication will be excluded from the study. Further, it was decided that individuals who smoke three refers of cannabis or more daily will be excluded. Thus to control for normality, subjects will be asked three questions concerning their neurological history, their history of medication, and whether they smoke three reefers or more a day of cannabis. In addition, as a further control for normality, it was decided that all subjects will be required to have been in their present employment for a period of at least 6 months. The rationale behind this was that adults who have kept employment for this period are less likely to be brain damaged.

6.8 The influence of depression on memory

Lezak (1983) warns that the influence of depression on memory should be considered. But, Gass and Russell (1986) failed to find performance deficits on two tests of memory[1] in depressed patients while these were evident in brain damaged patients. These findings led them to postulate that the depression which is often present in neurologically-impaired patients is unlikely to be connected to their memory problems. They stress the need for caution before ascribing memory deficits to depression, especially in the presence of brain damage or other factors to which memory dysfunction can be attributed. They suggest that "clinical lore" frequently connects memory impairment with depression (Gass & Russell, 1986, p.262). The failure to establish this connection in their own and other studies means

---

1. The two tests used were the Wechsler Adult Intelligence Scale Digit Span subscale and the Wechsler Memory Scale-Revised Logical Memory subtest, which measured short- and long-term memory respectively (Gass *et al.*, 1986).

that the 'lore' could "in part" be due to statements by depressed patients who periodically depreciate their memory abilities (Gass & Russell, 1986, p.262). In addition, the elimination of subjects taking any type of sedative or psychiatric medication (see 6.7: p.112) is likely to eliminate subjects who are seriously depressed. It was therefore decided not to control further for the possible influence of depression.

PART THREE: METHODOLOGICAL PROCEDURES

Part Three consists of two chapters which concern the actual methods applied in this research. The first chapter, Chapter 7, concerns the rationale for a BGRT scoring system and the presentation of this system, which is used in this study. Chapter 8 describes the subjects and the actual procedures followed during the testing and the statistical analyses to which the raw data are applied.

CHAPTER 7: THE RATIONALE AND OPERATIONALIZATION OF THE WEISS-LACKS BGRT SCORING SYSTEM

The method of combining Lacks's (1984) and Weiss's (1970) systems are briefly repeated. Requirements for a BGRT are then presented. Next, certain terminology is defined for precise clarification. Following this, Weiss's (1970) scoring criteria are presented again for ease of reference. Subsequently, the compatibility of each of Lacks's (1984) 12 criteria with Weiss's (1970) criteria are individually discussed. The specific way of scoring each design concludes the discussion of each item. Finally the criteria for a BGRT scoring system are provided and tabulated.

7.1 The combined Weiss-Lacks scoring method for the BGRTs

In section 4.5 (p.84) it was decided that in most instances Lacks's (1984) BGT-*c* system would be used to operationalize Weiss's (1970) BGRT system. Where Weiss's (1970) criteria are vague, Lacks's (1984) system would be used in most instances, to clarify them. It was also decided in the same section that, where Lacks (1984) would regard a design as sufficiently incorrect to award it a negative point, it would be rejected as insufficiently correct to score a positive point. In addition, it was stated that in developing a BGRT scoring system, focal issues surrounding adherence to the tenet of the gestalt and established research literature, would be incorporated.

7.2 Requirements for a BGRT scoring system

Since Lacks's (1984) system is intended to detect neurological impairment, the criterion of a system specifically designed to detect neurological impairment has been considered. In addition, it was suggested that it would be advantageous if a BGRT scoring system was compatible to the BGT-*c* scoring system. The decision to combine these two systems could render the BGT-*c*'s and BGRTs' scores compatible, and therefore meaningfully comparable, across the three BGTs. It was also decided that recalled designs would be scored according to a positive system and that each adequately recalled design would be allocated a single scoring point (see 4.5.2.2: p.87).

Other requirements for a recall scoring system were suggested (see 4.5: p.84), and must now be taken into consideration. A recall scoring system for this research in question should:

(i) retain the concept of the gestalt and therefore penalize fragmentation, lack of integration and incomplete designs;

(ii) regard rotations as errors;

(iii) be rapid to score and relatively quick to learn;

(iv) be adequately defined in sufficient detail to facilitate its replication;

(v) report scores obtained for the BGRTs, in a clear, precise manner;

(vi) have available normative data which are stratified according to age and educational levels.

In this discussion a number of factors must be borne in mind as already discussed (see 4.5: p.84). Firstly, as stated (see 4.5.3: p.89), allowances must be made for the fact that the designs are being recalled from memory and not copied. This must be balanced with an adequate degree of stringency (see 4.5.2.1: p.86). Secondly, Lacks's (1984) system has an organic 'cut-off' point of five errors but with the BGRT a single error can render a design as inadequately recalled. This must be balanced with an adequate degree of stringency. The above concepts will be considered continually in the decisions made below.

7.3 Terminology

The terminology used to describe the quality of the recalled BG designs is confusing. For a lucid description of a BGRT scoring system it is necessary first to define the parameters under discussion. For clarity then, in this research, Weiss's (1970) *definitions* for scoring the recall test will be used but the *terminology* will be changed. Specifically, Weiss (1970, p.143) uses the terms of "good" and "total" recall which he states were introduced by

Sharav (1968, in Weiss, 1970, p.143).[1] 'Good' recall refers to a more accurate reproduction of a design than does 'total' recall, making these nominations obscure (Weiss, 1970). A 'good' reproduction is a perfect design, or one containing no more than a "single slight distortion" while 'total' recall denotes a design recalled entirely or partially, regardless of distortions (Weiss, 1970, p.143). The only stipulation here is that a design should be sufficiently identifiable to differentiate it from a "fabulation" (Weiss, 1970, p.143). 'Good' does not adequately describe the stringency of the criteria which it represents. The term used by Rogers and Swenson (1975, p.919) of "near perfect" is more succinct and is congruent with the tenets of this BGRT scoring system. Thus, 'near perfect' will define the reproductions which are scored as correct, according to Weiss's (1970, p.143) criteria for 'good' recall.

7.4 Weiss's (1970) and Lacks's (1984) scoring criteria

7.4.1 Weiss's (1970) scoring criteria

Weiss's (1970, p.143) stipulations for near perfect recalled BG designs require that they have no distortions or only a "single, slight distortion". Therefore Weiss (1970, p.143) excludes the following errors:

a. Simplifications, i.e. the replacement of a complex (part) design, e.g. hexagon or square, by a simpler (part) design, e.g. diamond or triangle respectively; "swallowing-up" or "cutting-off" of a corner in Figs. 7 or 8 or shortening of Figs. 1 or 2 were not excluded.
b. "Extreme" i.e. edge-to-edge perseverations in contrast to "moderate" ones, where the number of units (dots or triplets) exceeds that of the design, but does not reach "extreme" proportions.
c. Rotations for more than 45°.

The corollary is that Weiss's (1970) system allows:

- a "slight distortion" (Weiss, 1970, p.143);
- "moderate" perseveration (Weiss, 1970, p.143);
- the failure to reproduce the angles on designs 7 and 8;
- rotations of less than 45°;
- the shortening of designs 1 and 2.

---

1. Weiss (1970) states that these terms were introduced by Sharav (1968, in Weiss, 1970) but this differentiation in the quality of the figures was in fact used in 1957 by Olin & Reznikoff with meanings analogous to those of Sharav. Presumably then, Sharav (1968) obtained these definitions from this earlier paper.

7.4.2 Lacks's (1984) scoring criteria

Lacks's (1984, pp.86-109) criteria are as follows (see Appendix 2 for details: p.232):

1. "Rotation" - 80° to 180° (all designs are scored);
2. "Overlapping" (designs 6 and 7 are scored);
3. "Simplification" (all designs are scored);
4. "Fragmentation" (all designs are scored);
5. "Retrogression" (all designs except 4 and 6 are scored);
6. "Perseveration":

   Type A (designs 2,3 and 5 only are scored);
   Type B (designs 1,2 and 3 are scored);
7. "Collision or collision tendency" (all designs are scored);
8. "Impotence" (all designs are scored);
9. "Closure" (designs A, 4, 7 and 8 are scored);
10. "Motor incoordination" (all designs are scored);
11. "Angulation" - (designs 2 and 3 only are scored);
12. "Cohesion" (all designs are scored).

7.5 Consideration of each scoring item

As discussed (see 4.4: 83) Weiss's (1970) criteria are considered insufficient for an objective scoring system, particularly if a high degree of validity and inter-scorer reliability is envisaged. Each of Lacks's (1984) 12 scoring items are considered below, along with Weiss's (1970) criteria.

7.5.1 Rotation

The significance of rotational errors in the identification of brain damage has already been emphasized (see 4.2.5: p.81) and it is once again stressed that rotations are neurological signs. Weiss (1970) specifies that a rotation of more than 45° should be scored as incorrect. Lacks's (1984, p.86) system contains two criteria regarding rotation errors: (i) rotation by 80° to 180° for all designs, and (ii) the "angulation" criterion is applicable when designs 2 and 3 are rotated between 45° and 80°. As already discussed, with Lacks's (1984) system, the number of errors is important while on the BGRT system a single error eliminates the design. Thus the combination of Lacks's (1984) two criteria into a single one, simplifies scoring procedures without fundamentally changing the system. For ease of scoring, then, all errors of rotation will be combined in one category under the term rotation. A slight discrepancy between the two systems exists. Lacks (1984) refers to angles of 45° or more, and Weiss (1970) to angles of more than 45°. This difference results in a 1° discrepancy which is unlikely to make a significant difference. For accuracy of scoring rotations of 45° or more will be regarded as an error.

Combining the above with Lacks's (1984) stipulations for rotation results in seven designs not being scored for rotations of between 45° and 80°. This leaves a discrepancy between the two systems, since Weiss (1970) advocates that *all* designs which are rotated more than 45°, are scored as incorrect. But, as already discussed, Lezak (1983) states that rotations are diagnostic signs (see 2.1.5: p.12). Further, on the BGT, research has established that rotations have serious neurological implications (see 4.2.5: p.81). Also, the stringency of this BGRT scoring system requires a near perfect reproduction. Thus, Weiss's (1970) criteria will be adhered to and it was decided that a rotation of 45° or more, for all of the nine designs, will constitute an inadequately recalled design.

7.5.2 Overlapping

Weiss (1970) makes no mention of such a criterion. However, an overlapping error, as defined by Lacks (1984) is clearly more than a 'slight' distortion. Therefore, Lacks's (1984) definition of overlapping (see Appendix 2: p.232), which is applicable to designs 6 and 7 only, will be applied to the BGRT scoring system. Where such an error occurs, the design will be regarded as inadequately recalled.

7.5.3 Simplification

Both Weiss (1970) and Lacks (1984) consider simplification in their scoring systems. Lacks (1984) scores simplification errors for all designs. In addition, both authors make allowances for failure to reproduce angles accurately. Weiss (1970, p.143) states that for designs 7 and 8, the "swallowing-up" or "cutting-off" of an angle is allowed. Lacks (1984, p.91), in her criterion of simplification, provides an illustration of design 8 and states that "curves" which are "substituted for angles" should not be scored as incorrect. However, it is not clear whether this applies solely to design 8 or to all the designs. Attempts to trace Lacks to clarify this, have been unsuccessful to date. If angles were substituted with curves in all the designs, too many inaccuracies would be allowed. For example, the diamond on design A, would result in a circle. Clearly, allowances for angles on all designs do not appear logical. However, designs 7 and 8 are comparatively complex to *copy*. It would seem acceptable then, to allow curves for angles on these two designs.

Weiss (1970, p.143) refers to simplification as being a "simpler" design or shape than the original one. Lacks's (1984, p.90) criteria are more sophisticated in that she distinguishes between simplification which constitutes a design drawn "in a simplified or easier form *that is not more primitive, from a maturational point of view*" (*italics added*). Designs which are reproduced in a more primitive way fall under her category of "retrogression" (Lacks, 1984, p.94). Both simplification and retrogression constitute a type of simpler form, with the distinction being as to whether one is more primitive than the original. Both can be regarded as more than a minor error. It is therefore in keeping with Weiss's (1970) system to retain both categories. However, for ease of scoring, since errors which the scorer must identify for these two items are quite different, they will be kept separate. Included in Weiss's (1970) specifications of simplification is the shortening of designs 1 and 2. These will be considered under Lacks's (1984) criterion of 'fragmentation' which follows. Simplification will disqualify a design as correctly recalled, if it fits Lacks's (1984) description.

7.5.4 Fragmentation

This criterion refers to the breaking up of the gestalt. As already discussed (see 4.2.4: p.80), the preservation of the gestalt is fundamental to the tenets of the test. If the pattern organization is disturbed by the gestalt being fragmented, then disorganization in thought processes is occurring, according to Bender (1938). Thus, it is logical that an error on this item can be considered more than a minor distortion. Lacks (1984) stipulates that fragmentation requires 6 or fewer dots for design 1 or 6 or fewer rows of dots for design 2. However, Weiss (1970) permits the shortening of designs 1 and 2 with no further definitions. A moderate, but not an extreme shortening of the designs will be allowed, considering that this is a memory test. It was therefore decided that a 50% or greater shortening of the designs would not be allowed, which renders both systems identical. For design 1, which has 12 dots, this means 6 or fewer dots will be disqualified as incorrect. Design 2 has 11 rows of ovals so again 6 or fewer would constitute an extreme distortion.

7.5.5 Retrogression

The criterion of retrogression was discussed under section 7.5 on simplification. Further, Bender (1938) notes that regression occurs as perceptual processes become impaired. This item will therefore be scored as specified by Lacks (1984), with designs which would score an error on the BGT-*c* being regarded as inadequately recalled on the BGRT(s).

7.5.6 Perseveration

Weiss (1970, p.143) allows "moderate" perseveration and disallows "extreme" or "edge-to-edge" perseveration. The problem with this definition is that some subjects reproduce edge-to-edge designs in a design without perseveration in that the reproduction of the design is evenly distributed across the full width of the page with the correct number of dots or rows of circles. In contrast, other reproductions can be extremely perseverated but are not edge-to-edge as they are drawn very small. Lacks (1984, p.245) defines this as "Type B" perseveration and uses the definition of more than 14 dots on design 1 and more than 13 rows of ovals on design 2. It was decided that 'moderate' perseveration would be operationalized as less than 50% of the original design, i.e. 16 or fewer dots for design 1 and 17 or fewer rows of ovals for design 2. Thus extreme perseveration constitutes 17 or more dots for design 1 and 18 or more rows of ovals for design 2. A 50% perseveration of design 3 would constitute two additional lines of dots.

Lacks (1984, p.245) has a second class of perseveration criteria referred by her as "Type A" which are applied to the carrying over of a response from a previous design, i.e., continuation of the previous design. For example, the continuation of the dots from design 1 by substituting the ovals in design 2 with dots instead. As already discussed, Lezak (1983) regards perseveration as a strong neurological indicator even on its own (see 2.1.5: p.12) and Bender (1938) strongly indicates perseveration as problematic. Thus this criterion will be retained for the same designs to which Lacks (1984) applies it, and where evident a recalled design will be scored as incorrect.

7.5.7 Collision or Collision Tendency

This criterion applies to the spatial proximity of the designs to one another, which involves spatial organization. The initial designs are on individual cards. However on the BGRT, the designs are being 'drawn' from images in the subject's mind. Clarity of organization is already being tested as neither parts of a design nor mixed designs are marked as correct. Since the subject no longer has the visuospatial advantage of viewing each design, it was decided not to score this item as a BGRT error.

7.5.8 Impotence

There are two problems with the scoring of this item on a BGRT; (i) the dynamics of recall are different from that of copying, and (ii) it is a difficult item to observe on the BGRT(s), and especially so in a group context. Most important, on the BGT-*c*, the design is in full view and subjects who reproduce the design incorrectly despite repeated rubbings out and new attempts, are experiencing difficulties with some aspect of the visuomotor perceptual process. In contrast, on the BGRT(s), the subjects are 'copying' from images in their minds. Once the design is drawn, subjects might compare it with the image in their minds and find it lacking. Thus rubbing out could be a correction and not impotence. The second, minor, consideration is that impotence is difficult to observe during the BGRT(s), particularly in a group context. With the BGT-*c*, it is easy to discretely observe a subject who continually rubs out and/or expresses frustration, and who does not turn the card, clearly indicating he or she is still working on the same design. With the BGRT(s) there is no card to be turned. This makes it difficult to establish whether the subject is rubbing out the same design or whether he or she has proceeded to the next design unless the tester obtrusively stands where he or she can observe what the subject is drawing on the paper. The subject could get anxious if the tester observes in such a way for a duration long enough to recognize impotence. Because of the different processes involved and the secondary problem of observation it was decided to eliminate this criterion.

7.5.9 Closure

Closure is another concept which is also related to the gestalt function. The error of closure effectively means breaking the gestalt, the significance of which has already been discussed (see 4.2.4: p.80). Like fragmentation, this item will be scored precisely according to Lacks's (1984) definition, and if a design would be awarded an error score on the BGT-*c*, it will be regarded as inadequately recalled.

7.5.10 Motor incoordination

Motor incoordination is either present or not in an individual's handwriting and its presence or absence is likely to be fairly constant. Therefore where motor incoordination exists, it is likely to be present in more than a single design. Also, it could indicate a muscular, as opposed to a perceptually-based visuomotor function and therefore alone, it is not necessarily indicative of brain dysfunction. On the BGT-*c*, motor incoordination is regarded as a single error only, even if it is evident in each design. Since the BGRT is scored differently as already discussed (see 4.5.2.1: p.86), if motor incoordination is scored and occurs in most of the designs, the subject is being disadvantaged more than once for a single error that is not necessarily due to brain dysfunction. For this reasoning, it was decided to omit this error from the scoring.

7.5.11 Angulation

Part of Lacks's (1984) criterion for this scoring item has been included in the rotation category (see 7.5.1: p.118) for the BGRT scoring system. Two further criteria were not included - failure to copy designs 2 and 3 with any angulation at all and variance of at least 50° of angulation in the rows of circles for design 2. A change in the angulation of an entire original design of more than 45° has been covered under rotation. It is important to note that even on the BGT-*c*, Lacks (1984) states that the angulation in design 3 is particularly difficult to reproduce. Therefore only the failure angulation on this design will disqualify the recalled design.

7.5.12 Cohesion

The rationale in considering cohesion is similar to the criterion of collision. If the proportions between two parts of a single design or between two different designs are distorted while the design is visible to the subject, then perception is distorted in some way. However, this is not necessarily true if the subject is recalling from memory. It is also sufficient to require the recall of a design after copying it once. To expect a subject to remember the actual scale of the design is considered too demanding for a recall test. Also, Lacks (1984) warns that this is an error which is too often overscored. Therefore, for similar reasoning to that discussed under collision, this criterion will be excluded from the BGRT scoring system. The next section will summarize and tabulate the decisions made above.

7.6 The operationalization of the Weiss-Lacks scoring method for the BGRTs

7.6.1 A recall scoring system

The final recall scoring system has eight scoring points. Each design which meets the 'near perfect' criteria, will be awarded a single point. From the rationale presented in Section 7.1 (p.115) the following criteria applicable to the designs, indicated by number, were established for a BGRT scoring system:

Criteria		Designs scored
1.	Rotation - 45% or more scored as incorrect	all
2.	Overlapping	6,7
3.	Simplification	all
4.	Fragmentation	all
	design 1: fewer than 6 dots design 2: fewer than 3 rows of ovals	
5.	Retrogression	all except 4 and 6
6.	Perseveration	
	type A - as per Lacks (1984)	2,3,5
	type B -	1,2,3
	design 1: 17 dots or more are incorrect design 2: 18 or more rows of ovals are incorrect design 3: 2 or more extra rows of dots are incorrect	
7.	Not scored for recall	-
8.	Not scored for recall	-
9.	Closure	A,4,7,8
10.	Not scored for recall	-
11.	Angulation - falls into rotation category except for no angulation at all on both designs and a 50% variance of angulation for design 3 only	2,3
12.	Not scored for recall are incorrect	-

The scoring criteria which apply to the BGRT system can now be applied to each design to facilitate scoring, as follows:

7.6.2 Individual criteria as they are applied to each BG design

Criterion

A.	Rotation, simplification, fragmentation, retrogression, closure
1.	Rotation, simplification, fragmentation, retrogression, perseveration - type B
2.	Rotation, simplification, fragmentation, retrogression, perseveration - type A and B, angulation
3.	rotation, simplification, fragmentation, retrogression, perseveration - type A and B, angulation
4.	rotation, simplification, fragmentation, closure
5.	rotation, simplification, fragmentation, retrogression, perseveration - type A
6.	Rotation, overlapping, simplification, fragmentation
7.	rotation, overlapping, simplification, fragmentation, retrogression, closure
8.	Rotation, simplification, fragmentation, retrogression, closure

The final scoring criteria have been tabulated below for rapid and easy reference when scoring the BGRTs (see Table 1: p.126). It is important to note that since Lacks's (1984) BGT-*c* system has been used to operationalize Weiss's (1970) system, it presupposes a thorough knowledge and understanding of the Lacks (1984) system. For this reason, Lacks's (1984) system has been included in this thesis in full (see Appendix 2, p.232). Once Lacks's (1984) system is learned, this BGRT system will be clear and easy to use.

Preliminary normative data which are stratified for education, for adults with less than 9 years education using this BGRT scoring system are presented in this research (see Table 4: p.138). Further research with larger sample numbers is necessary to obtain normative data which can be stratified for age and education.

Table 1: A scoring system for the BGRTs

Design	1 Rotation >45°	2 Over-lapping	3 Simpli-fication	4 Fragmen-tation	5 Retro-gression	6 Perseveration A	6 Perseveration B	7 Closure	8 Angulation
A	Score	-	Score	Score	Score	-	-	Score	-
1	Score	-		6 or less dots incorrect	Score	-	More than 17 rows of dots incorrect	-	-
2	Score	-		6 or less rows of ovals incorrect	Score	Dots for ovals incorrect	18 rows of ovals or more incorrect	-	50% variance or no angulation incorrect
3	Score	-	Score	Score	Score	Ovals for dots incorrect	2 or more extra rows of dots incorrect	-	Only no angulation incorrect
4	Score	-	Score	Score	-	-	-	Score	-
5	Score	-	Score	Score	Score	Circles for dots incorrect	-	-	-
6	Score	Score	Score	Score	-	-	-	-	-
7	Score	Score	Score curves	Score	Score	-	-	Score	-
8	Score	-	Score (curves)	Score	Score	-	-	Score	-

CHAPTER 8: METHOD

The problems and uncertainties regarding scoring of the expanded BGT have been clarified and the rationale and manner of the administration as well as the selection of subjects have been decided upon. Thus all the necessary decisions for conducting the research have been made. The actual method of the research can now be presented. Consequently, this chapter describes the research itself in four subsections. A description of the subjects who participated in this research is followed by a brief description of the tests administered. Then the procedure of administration is presented. Finally this chapter concludes with a note on the statistical procedures applied to the data obtained. This ends Part Three.

8.1 Subjects

Subjects were recruited through a dual system of contact with their employers and the various unions attached to these organizations. The method of recruiting subjects will be further dealt with in the discussion (see 10.1.1: p.152). Participation was voluntary and strict measures were taken to ensure this, considering the mode of recruiting.

The testing process was begun with 181 adults resident in the Grahamstown district. However, for reasons discussed below, 41 subjects did not complete all the tests. A further 58 subjects were excluded from the preliminary normative data for low-educated adults. Of these, 5 were beyond the age range decided upon i.e. 60 years and older, 21 reported possible neurological damage and 13 were beyond the education limit of less than 9 years of education. In all, 82 subjects met the restrictions of the category for normative data.

However, data from 121 subjects were submitted to statistical analysis. In addition to the non-clinical sample of 82 subjects, it was decided to investigate the scores of another 39 subjects for the following reasons. Due to the large number of subjects (21) eliminated from the sample for possible neurological damage, it was decided to investigate their scores as well. In addition, 13 subjects with education levels ranging from 9 to 11 years of education were used in this research.[1] Although this is beyond the educational level under

1. Shop stewards with higher levels of education participated in order to demonstrate their willingness to participate, both to the researcher and their union members. In some instances, where requested by unions, the level of education was deliberately not stipulated when recruiting volunteers (see Appendix 6B, p.260).

consideration, these scores were used for comparison purposes and for cross-validation of results. Information from the 5 subjects of 60 years old and older is also briefly considered. All these subjects had been in their current employment for at least 6 months. None of the subjects reported difficulty in seeing the BG designs, although due to the detailed briefing some individuals declined to participate in the research because of poor eyesight.

The remaining 60 subjects were totally eliminated from the research for the following reasons: 3 individuals had 12 years of education; 7 subjects were excluded as they reported smoking three reefers or more of cannabis daily, 5 individuals reported sustaining head injuries *and* smoking three reefers or more of cannabis a day; 4 subjects reported they were currently taking sedative medication; 1 person refused to continue and was also highly anxious; 6 individuals showed high levels of anxiety and testing was discontinued; 1 subject went to get some water and did not return; 23 subjects did not complete the entire test;[2] and in one instance the researcher terminated testing of 10 subjects because the situation was inappropriate as the employer intimidated his staff during the testing.[3] Of the 181 subjects, 41 reported having sustained at least one head injury with unconsciousness, which constitutes 22,65% of the sample.

Subjects used to establish preliminary normative data all reported that they had not sustained any head wounds that had led to any period of unconsciousness. They stated they had not suffered from neurological diseases[4] such as epilepsy, meningitis, encephalitis or hydroencephaly. The subjects also denied smoking three or more reefers of cannabis a day, that they had taken any psychiatric medications or were taking sedatives at the time of testing. The sample ranged in age from 22 to 58 with education levels varying from no formal education at all to less than nine years, i.e. an uncompleted standard seven.

---

2. In some instances testing was not completed due to time constraints and other problems (see 10.1.3: p.155).

3. This will be discussed further (see 10.1.2.1: p.154).

4. This aspect of the control for normality will be fully dealt with under the discussion (see 10.1.3.2: p.156 and 10.1.3.3: p.157).

8.2 Measures

8.2.1 The BGT-*c*

The nine BG designs from Bender's Test Manual (1946), pasted onto cards 15 cm x 10 cm and correctly orientated, are placed parallel to the top of each subject's paper. Subjects are required to copy each design. There is no time limit, but the time for completion is unobtrusively noted. Protocols were scored by the author according to the Lacks (1984) BGT-*c* (see Appendix 2, p.232).

8.2.2 The BGRT-*i*

Immediately the BGT-*c* is completed, the copied protocols and cards are removed. Subjects are asked to erect a screen around themselves,[5] and are then requested to draw as many of the designs as they can remember. No warning of the BGRT-*i* is given nor are subjects encouraged to remember more designs. The Weiss-Lacks scoring method developed by this author (see Chapter 7: p.115), was used to score this test.

8.2.3 Goodenough's "Draw-A-Man" Test

This test is used in this research to provide a delay with interference between the immediate and delayed recall versions of the BGT. According to Goodenough's instructions (Harris, 1969), subjects are asked to "draw the best man that you can".

8.2.4 The BGRT-*d*

This is administered and scored in the same manner as the BGRT-*i*. It commences 10 minutes after the completion of the BGRT-*i*.

---

5. The reason for this is explained later (see 10.1.5.3: p.159).

8.3 Procedure

Since one of the aims of this study is to standardize the administration of the expanded BGT in a manner which can be replicated (see Chapter 1: p.1), the procedure of testing is described in detail below.

The subjects were mostly tested within their own environment and where possible this researcher visited the location beforehand to assess its suitability. Where no facilities for testing at the workplace of the subjects existed, subjects were tested in an environment that was considered unthreatening to them by the researcher. Attempts were made to ensure that the testing environment was quiet.[6] Subjects were given soft pencils and erasers. Subjects were also given their own pile of A4 paper but were instructed not to press on the whole pile. This was requested in order to avoid the imprints of the page being written on, appearing on subsequent pages.

Testing was conducted in groups, administered either by the author alone, or together with a research assistant. When the researcher was working alone, up to 8 individuals were tested simultaneously and a maximum of 14 subjects were tested in each session with the aid of a research assistant. The research assistant was a psychology student, fluent in Xhosa, who was well briefed regarding the method of testing prior to the test administration. He had assisted in previous research with the same tests. For ease of reference, both individuals will be referred to as researchers in the discussion below. In all, 22 groups were tested. Since the testing was timed to suit the employer or subjects giving up their free time, most often groups smaller than the maximum suggested, were tested. Where the researcher worked alone, a Xhosa translator was used. Some groups preferred to have their own translator while others had no preference and an outside translator was used. The translators were briefed beforehand not to alter the instructions. Where translation into Afrikaans was necessary, the researcher did this herself as she is fluent in this language.

Where space permitted, the subjects were spaced well apart to avoid anyone busy with one of the BGRTs seeing the cards of a person who was still on the BGT-*c*. Where such spacing

---

6. This will be discussed further in the discussion (see 10.1.4: p.?).

was not possible, subjects were asked to erect a screen around themselves on the completion of the BGT-*c*.

Prior to the commencement of testing, subjects were asked whether they understood the purpose of the test and whether they had any questions. Testing did not begin until all the subjects' concerns were addressed. All groups were assured that their results would remain confidential, especially from their employers. Subjects were then given a chance to withdraw from testing if they so wished.

It was then demonstrated that the pack of cards each subject would receive must remain parallel to the top of the paper. Subjects were warned that the researcher might draw a line while they were working (this was demonstrated before testing began), but that this should not concern them. This method of marking rotations of the card or paper was decided upon earlier in this thesis (see 4.2.5: p.81). Subjects were instructed not to use mechanical aids such as coins and rulers but were told that the use of the erasers provided was permitted.

Subjects were each given a number which they were asked to place in front of them.[7] Subjects were then requested to write their names, their highest completed year of education and age or date of birth on their piece of paper. Those with an incomplete standard seven where asked to indicate this specifically. Three questions were then asked to control for normality (see 6.7: p.112) and subjects were asked to place a cross for 'no' and a tick for 'yes' in a vertical line. This was demonstrated. Those with questions or problems were asked to raise their hands.[8] When this was completed, subjects were asked to turn their paper over, so that they each had a blank sheet in front of them.[9]

Just before the administration of the tests began, individual packs of (15 cm x 10 cm) cards with the BG designs were placed parallel to the top of each subject's paper. As direction

---

7. This was to facilitate group observation and administration and is fully described in Appendix 6A. (p.260).

8. This was to mask those who did not know their age or who could not write their name. The reason and the procedure is more fully described in Appendix 6B (p.260).

9. This was to facilitate marking protocols blind, without knowledge of educational level nor possible neurological impairment.

and sequence is fundamental (see 5.2.2: p.97), each pack was checked prior to each testing. Subjects were instructed to copy the designs 'as best you can' and 'just as you see them'. The subjects were allowed to angle the paper provided that the cards remained parallel to the top of their page. Where a card was reorientated, the researcher(s) indicated this non-verbally by placing the cards back in the original position. Where subjects insisted on rotating BG designs or paper, the researcher(s) marked the baseline angle on the protocols. The cards were removed after completion of the BGT-*c*.

As each subject completed the BGT-*c* at his or her own pace, the BGRT-*i* was then administered, followed by Goodenough's 'Draw-A-Man' Test and after a 10 minute interval, the BGRT-*d* was administered. Subjects were told this was the last task.[10] Opportunity to complete the Goodenough drawings was then provided for those who had not been able to complete the Goodenough drawings in the specified time. At the beginning of each BGRT the researcher(s) removed any paper on which designs had been imprinted.

The researcher(s) were alert and attentive during the entire testing process and noted Lacks's (1984) exclusionary observational criteria plus the two additional ones introduced by this author (see Appendix 3: p.257). Completed protocols were marked with the subjects' numbers. As each protocol was set down on a table, the time was noted from a watch lying unobtrusively on the table. Observational notes and timing were recorded on the subject sheet (see Appendix 5: p.259) alongside the number each subject had placed above their sheets of paper.

Apart from thanking the subjects each time they completed a protocol, no comment was made about the number of designs recalled. Anxiety about not being able to recall more designs was counteracted with "it is only important that you have done your best". Questions asked by subjects during the BGRTs as to how many designs the original pack contained, were answered by informing the subject that this was not important as they merely needed to reproduce the designs they could recall, or similar reassurances.

Some subjects enjoyed the 'Draw-A-Man' task so much that they wanted to continue drawing rather than execute the BGRT-*d*. This problem was resolved by telling subjects that

---

10. The reason for this is discussed in section 10.1.3.4 (p.157).

when they finished they could draw other pictures. Thus, when subjects wished to, those who were finished were given more paper on which to draw pictures of their own choice. Where these drawings were offered to the researcher "to help with the studies" the pictures were accepted with thanks. Afterwards, subjects were debriefed[11] and questions were answered. They were thanked and again reassured that their performance was satisfactory and that the information given to the researcher was confidential.

The BGT-*c* was scored according to Lacks's (1984) system (see Appendix 2: p.232) and the BGRTs according to the system presented in this research (see Chapter 7: p.115). The three protocols of each subject were kept together in order to score rotations[12]. To facilitate rapid and accurate scoring on the BGT-*c*, a list of the 12 scoring criteria was stapled to each protocol (see Appendix 7: p.261). This was also useful for moderating scoring. To ensure a high degree of scoring reliability, the first 20 protocols which were scored (during which time this author reacquainted herself with Lacks's (1984) scoring system) were scored again at the end. In addition a random moderation of every 10 protocols was conducted by scoring each one again, blind. In all but one instance, the variance in the scoring did not exceed one point. This author found that it was both easier and more reliable to score the protocols of the BGT-*c* and each BGRT separately. This is preferable for the validity and reliability of the scoring in order to score blind. Being aware of the performance on one test could well influence the researcher when evaluating another.

8.4 Statistical analysis

No extreme scores were excluded from the non-clinical sample as these could represent the range of BGT performance. Descriptive statistics comprising the means, standard deviations and ranges were established for the normative data subject group for each of the four education levels on the BGT-*c*, the BGRT-*i* and the BGRT-*d*. In addition the same statistics were also computed for the time taken to complete each test and the overall time for the four tests were also computed. To test for differences between the expanded BGT scores and performance time across the educational categories, one-way analyses of variance tests (ANOVAs) were computed. Where significant F-ratios were obtained, Tukey

---

11. The need for debriefing is elucidated in the discussion (see 10.1.3: p.155).

12. The reason for this is explained in the discussion (see 10.1.7: p.161).

multiple range tests were performed, to indicate where the differences lay. The above statistical procedures were repeated with the possibly brain damaged group and the group with 9 or less years education (see 8.1: p.127). In addition, to test for mean differences between the scores and performance time of these two groups, one-way ANOVAs were run.

Multiple regression analyses were run to establish the correlation of relationships between certain data. Analyses were performed between each test for the sample as a single group, as well as for each education level separately. The mean scores of subjects in this study were compared with the mean scores of subjects in other studies, using multiple *t*-tests. Since the subjects were not randomly selected, as it was ethically necessary to recruit subjects on a voluntary basis (see 10.1.8.1: p.163), population parameters such as equality of variance and normal distribution could not be assumed. Thus, where applicable, the data were submitted to Levene's test for equality of variance. Finally, the computational formulae presented later (see 10.4: p.196), to obtain a single BGT quotient, were also applied to the scores of the expanded BGT.

PART FOUR: PRESENTATION AND EVALUATION OF RESEARCH FINDINGS

Part Four brings this study to an end. The results obtained from this research are presented in Chapter 9. Chapter 10 then discusses issues of interest to the study and includes explanations of the statistical data as well as non-statistical information which emerged during the course of the study. Finally, Chapter 11 concludes this research, summarizing the significant findings and highlighting new research questions which have emerged in the course of this research.

CHAPTER 9: RESULTS

9.1 Demographic data

The mean educational levels, ages and genders of the non-clinical subjects used in this study appear in Table 2 below. Table 3 below provides the same data for subjects with possible neurological impairment.

9.1.1 Age

A one-way ANOVA revealed a significant difference in age between the educational levels (F =5,15, df =90, p < 0,0001). The 0 years of education group is older than the 7-9 and 9-11 years of education groups. A Tukey's multiple range analysis found this age difference to be significant at a 1% level. The 1-3 years of education group is also older than the 7-9 and 9-11 years of education group, with a minor significance at the 10% level. No other significant age differences existed. For the possibly brain damaged group as a whole, the mean age is 39,48, as presented in Table 3. These data are close to that of the sub-total of the group with 0-<9 years education, which is 39,85 years as shown in Table 2.

9.1.2 Gender

One-way ANOVAs consistently revealed that gender effects were non-significant on the scores of the expanded BGT for the sample as a whole (BGT-*c*: F =0,69, p =0,408;

BGRT-*i*: *F* =0,03, *p* =0,895; BGRT-*d*: *F* =1,17, *p* =0,283). One-way ANOVAs of the three BGTs across the five educational levels also produced non-significant F values, except on the BGRT-*d*, for the 1-3 years of education group. However, the number of females was too small, as can be seen from Tables 2 and 3, for the analyses of the individual educational levels to be statistically sound. In addition one-way ANOVAs revealed no significant influence of gender on performance time for the group as a whole, for any of the BGTs. Levene's tests for equality of variance were non-significant in all instances. Thus, gender is unlikely to be responsible for differences in scores or performance time on the expanded BGT.

Table 2: Demographic details of subjects used to establish preliminary normative data

Educational level (years)	*n*	Years of Education			Age			Gender	
		M	*SD*	Range	*M*	*SD*	Range	*M*	*F*
0	15	0	-	-	45,79	5,62	39-55	12	3
1 - 3	15	2,27	0,80	1-3	43,33	11,72	22-57	13	2
4 - 6	19	5,32	0,75	4-6	40,05	9,08	26-57	16	3
7 - <9	33	7,90	0,49	7-8,5	35,45	9,63	22-58	28	5
Sub-total	**82**	**4,84**	**3,13**	**0-<9**	**39,85**	**10,04**	**22-58**	**69**	**13**
9 - 11	13	10,00	0,91	9-11	34,15	7,51	24-47	10	3
TOTAL	**95**	**5,72**	**3,65**	**0-11**	**39,07**	**9,9**	**22-58**	**79**	**16**

NOTE: The 9-11 years of education group was beyond the education range established for this research. However, this information is referred to for comparison and is therefore included in this table.

Table 3: Demographic data of subjects with possible neurological impairment

Educational level (years)	*n*	Years of Education			Age			Gender	
		M	*SD*	Range	*M*	*SD*	Range	*M*	*F*
0 - 9	21	4,33	3,35	0 - 9	39,48	10,24	24-56	20	1

9.2 Performance on the expanded BGT

9.2.1 Educational effects on the expanded BGT

Descriptive statistics comprising of the means, standard deviations and ranges for performance on the expanded BGT across five educational levels appear below in Table 4. The same data for the group with possible neurological impairment appears in Figure 5 below.

Table 4 indicates that the scores of the BGT-*c* improve as the educational level increases. The only exception is an anomaly of the scores of the 0 years of education group for the BGT-*c* and BGRT-*i*. These are higher than the mean scores of those with 1-3 years and 4-6 years of education. A further noteworthy deviation from the trends indicated by this data is the range of scores for the 1-3 years of education group. The maximum score obtained by this group on the BGT-*c* is -4, which is only one point below Lacks's (1994) neurological 'cut-off' score of -5. The minimum score is -9. This contrasts to the highest score of -2 and the lowest score of -7 for the entire remaining sample. This will be addressed in the discussion (see 10.2.2: p.166).

The trend of scores improving as educational level increases is also found on the BGRTs, and the anomaly of mean scores for the 0 years of education group disappears on the BGRT-*d*. The range of the number of designs recalled also increases as the level of education does, except for the BGRT-*i* for the group with 9-11 years of education. This is below the recall of the previous educational group and can be expected as it is a smaller group which was not initially intended for inclusion in this sample, as already discussed (see 8.1: p.127).

A one-way ANOVA found a significant difference at the 1% level across the educational groups for all three tests. On the BGT-*c* F=8,72 (df=4, p<0,0001), on the BGRT-*i* F=6,61; (df=4; p<0,0001), on the BGRT-*d* F=7,02; (df=4; p<0,0001) and for the sums of the means of the two BGRTs F=7,02 (df=4; p<0,0001). Tukey's multiple range analysis tests indicate the significant differences between educational levels for each of the three BGTs as presented in Table 6. In addition, a one-way ANOVA found a significant difference between the 0-<9 year educational group as a whole and the group with 9-11 years of education (F=5,6; df=1; p=0,02). Levene's tests for equality of variance were non-significant in all cases.

Table 4: Means, standard deviations and ranges for performance on the expanded BGT by educational level of subjects used to establish preliminary normative data

Educational level (years)	*n*	BGT-*c*			BGRT-*i*			BGRT-*d*		
		M	*SD*	Range	*M*	*SD*	Range	*M*	*SD*	Ran ge
0	15	-3,73	1,49	-1 - -7	2,28	1,16	0 - 4	1,20	1,15	0 - 3
1 - 3	15	-5,53	1,60	-4 - -9	1,87	1,41	0 - 5	1,47	1,51	0 - 5
4 - 6	19	-4,79	1,79	-2 - -7	2,05	1,40	0 - 5	1,74	1,70	0 - 6
7 - <9	33	-3,12	1,71	-1 - -7	3,79	2,13	0 - 9	3,36	2,20	0 - 8
Sub-total	**82**	**-4,06**	**1,89**	**-1 - -9**	**2,75**	**1,88**	**0 - 9**	**2,24**	**2,01**	**0 - 8**
9 - 11	13	-2,77	1,36	0 - -5	3,77	1,59	2 - 7	3,46	1,94	0 - 7
TOTAL	**95**	**3,88**	**1,87**	**0 - -9**	**2,89**	**1,87**	**0 - 9**	**2,41**	**2,03**	**0 - 8**

NOTE: The 9-11 years of education group was beyond the education range established for this research. However, this information is referred to for comparison and is therefore included in this table.

Table 5: Means, standard deviations and ranges for performance on the expanded BGT for subjects with possible neurological impairment

Educational level (years)			*n*	BGT-*c*			BGRT-*i*			BGRT-*d*		
M	*SD*	Range		*M*	*SD*	Range	*M*	*SD*	Range	*M*	*SD*	Range
4,33	3,35	0 - 9	21	-5,81	2,25	-1 - -10	2,05	2.48	0 - 8	1,65	2.38	0 - 7

Table 6: Tukey's multiple range analysis tests for performance across education levels on the expanded BGT

Educational level (years)	BGT-*c*	BGRT-*i*	BGRT-*d*	BGRT-*i* + BGRT-*d*
0 *vs* 1-3	*	NS	NS	NS
0 *vs* 4-6	NS	NS	NS	NS
0 *vs* 7-<9	NS	*	**	**
0 *vs* 9-11	NS	NS	*	*
1-3 *vs* 0	*	NS	NS	NS
1-3 *vs* 4-6	NS	NS	NS	NS
1-3 *vs* 7-<9	**	**	**	**
1-3 *vs* 9-11	**	*	*	*
4-6 *vs* 0	NS	NS	NS	NS
4-6 *vs* 1-3	NS	NS	NS	NS
4-6 *vs* 7-<9	**	**	*	**
4-6 *vs* 9-11	**	*	-	*
7-<9 *vs* 0	NS	*	**	**
7-<9 *vs* 1-3	**	**	*	**
7-<9 *vs* 4-6	**	**	*	**
7-<9 *vs* 9-11	NS	NS	NS	NS
9-11 *vs* 0	NS	NS	*	*
9-11 *vs* 1-3	**	*	*	*
9-11 *vs* 4-6	**	*	-	*
9-11 *vs* 7-<9	NS	NS	NS	NS

NS = Not significant
- = $p \leq 0,1$
* = $p \leq 0,05$
** = $p \leq 0,01$

9.2.2 A comparison of the scores on the expanded BGT of non-clinical subjects with those having possible brain damage

A *t*-test comparison between the scores of the possibly brain damaged group and the normal group showed statistical significance at the 1% level for the BGT-*c* (t=3,64; df=93, p<0,001). However, the differences of the scores of BGRT-*i* and the BGRT-*d* are not significant, although the differences of the scores of the BGRT-*i* and the sum of the BGRTs between the two groups approach the 5% level of significance. Also noteworthy, is that the BGT-*c* mean of the possibly brain damaged group is -5,81, which falls within Lacks's (1984) neurological cut-off range of -5. The lack of statistically significant differences between the two groups on the BGRTs may be partly due to the substantial differences in sample numbers, with 82 subjects in the non-clinical group and 21 subjects in the possibly brain damaged group. Nevertheless, as Figure 2 shows, the mean scores for the possibly brain damaged group are consistently below those of the non-clinical group. The scores of the expanded BGT are represented in percentages to accord equal weight to each test. This process is explained later (see 10.4.1: p.196 and 10.4.2: p.197).

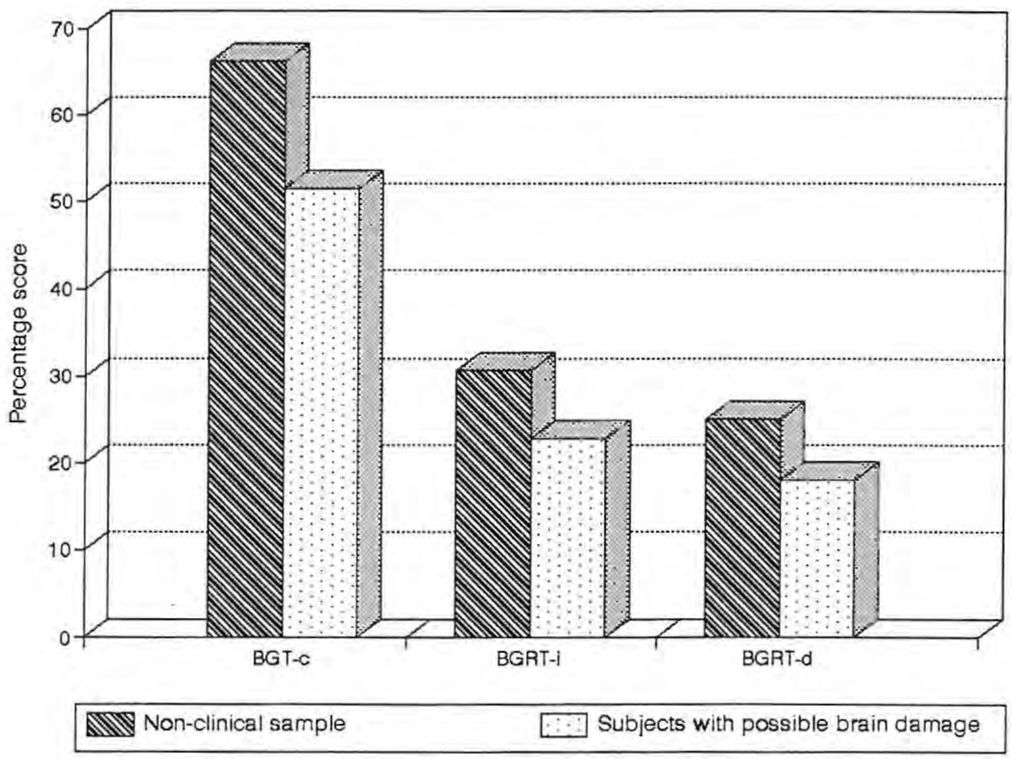


Figure 2: A comparison of the scores on the expanded BGT of non-clinical subjects with those having possible brain damage

As already discussed (see 9.1.1: p.135), the mean age of the possibly brain damaged group and that of the 0-<9 years of education group do not differ substantially. The educational range is identical and the mean years of education for the two groups are similar. The non-clinical sample has a mean and the educational level of 4,84 years (see Table 2: p.136), while the possible brain damaged sample has a mean educational level of 4,33 years as presented in Table 3 (p.136).

9.2.3 Lacks's (1984) neurological cut-off scores applied to the percentiles of BGT-*c* scores in this study

The percentiles of the scores for the BGT-*c* for the sample with 0-<9 years of education as a whole and for the group with possible brain damage for the same educational levels, are presented in Table 7 below.

Table 7: Percentiles of BGT-*c* scores for non-clinical subjects (above) and subjects with possible brain damage (below) both of which have 0-<9 years of education*

	Score	*n*	% of sample	Cumulative %
	-1	7	8,5	100,0
	-2	13	15,9	91,5
	-3	12	14,6	75,6
Lacks's cut-off → point	-4	16	19,5	61,0
	-5	18	22,0	41,5
	-6	6	7,3	19,5
	-7	7	8,5	12,2
	-8	2	2,4	3,7
	-9	1	1,2	1,2
TOTAL		82	100,0	

	Score	*n*	% of sample	Cumulative %
	-1	1	4,8	100,0
	2	1	4,8	95,2
	-3	1	4,8	90,5
Lacks's cut-off → point	-4	3	14,3	85,7
	-5	2	9,5	71,4
	-6	4	19,0	61,9
	-7	5	23,8	42,9
	-8	2	9,5	19,0
	-9	1	4,8	9,5
	-10	1	4,8	4,8
TOTAL		21		

* The non-clinical group has a mean educational level of 4,84 years and the possibly brain damaged group has a mean of 4,33 years of education

Where Lacks's (1984) neurological cut-off score of -5 is applied to this sample, 41,5% of the non-clinical sample 71,4% of the sample with suspected brain damage fall within the organic range. However, when the neurological cut-off score is increased to -6, only 19,5% of the normal sample falls within the neurological range. By contrast, a relatively high percentage of the sample with possible brain damage, namely 61,9%, remains within the organic range at a cut-off level of -6.

9.3 Correlations between the three tests on the expanded BGT

Multiple regression analyses established the relationship of the scores of the BGT-*c* with the BGRT-*i*, the BGRT-*d* and the sum of the two BGRTs, for the group as a whole, and for the individual educational levels. These revealed relatively high, significant correlations as shown in Tables 8 and 9.

Table 8: Multiple regression analyses indicating the correlations between the BGT-*c* and the two BGRTs for the non-clinical group as a whole

BGT-*c* and BGRT-*i*			BGT-*c* and BGRT-*d*			BGT-*c* + BGRT-*i* and BGRT-*d*		
r	*df*	*p*	*r*	*df*	*p*	*r*	*df*	*p*
0,65	91	<0,0001***	0,59	91	<0,0001***	0,64	91	<0,0001***

*** = p<0,001

Table 9: Multiple regression analyses indicating the correlations between the BGT-*c* and the two BGRTs individually, and between the sum of the two BGRTs across the educational levels

Educational level (years)	BGT-*c* and BGRT-*i*			BGT-*c* and BGRT-*d*			BGT-*c* and BGRT-*i* + BGRT-*d*		
	r	*df*	*p*	*r*	*df*	*p*	*r*	*df*	*p*
0	0,45	13	0,091	0,18	13	0,53	0,38	13	0,16
1 - 3	0,49	11	0,092	0,30	11	0,32	0,41	11	0,16
4 - 6	0,68	17	0,002**	0,58	17	0,009**	0,65	17	0,002**
7 - <9	0,58	31	0,0004**	0,64	31	0,0001***	0,63	31	<0,0001***
9 - 11	0,64	11	0,018*	0,55	11	0,049*	0,65	11	0,016*

* = p ≤ 0,05
** = p ≤ 0,01
*** = p ≤ 0,001

Similar correlations were also found for the group with possible brain damage, with the correlation between the BGT-*c* and the sum of the BGRTs showing significance at the 1% level (*r* =0,65; p<0001). The similarity in correlations for different groups supports the findings of a significant correlation.

For the above analyses, positive BGT-*c* standard scores were used, the conversion of which is explained later (see 10.4: p.196). The reason for using standard scores was to avoid dealing with inverse relationships, which complicates a direct comparison. Since Lacks's (1984) raw scores are negative and the BGRTs scores are positive, the comparison of raw scores would result in inverse relationships.

In addition, a *t*-test revealed differences between the scores of the BGRT-*i* and the BGRT-*d* at a 1% level of significance (*t* =4,98; *p* =0,002).

9.4 The influence of variables on the performance time of the expanded BGT

9.4.1 Performance time on the expanded BGT

The means, standard deviations and ranges of the time taken to complete the expanded BGT were computed across the five educational levels. These appear in Table 10 below.

Table 10: Means, standard deviations and ranges for performance time on the expanded BGT by educational level (time score in minutes)

Educational level (years)	*n*	BGT-*c*			BGT-*i*			BGT-*d*		
		M	*SD*	Range	*M*	*SD*	Range	*M*	*SD*	Range
0	15	25,33	9,56	14 - 45	12,93	4,48	5 - 24	10,00	3,55	4 - 18
1 - 3	15	22,84	6,41	12 - 36	14,62	2,84	10 - 22	9,46	2,73	5 - 15
4 - 6	19	18,63	7,98	7 - 35	10,90	3,50	5 - 19	8,53	3,38	4 - 15
7 - <9	33	17,00	6,43	6 - 40	10,33	3,51	5 - 18	7,94	2,87	2 - 15
Sub-total	**82**	**19,89**	**8,04**	**6 - 45**	**11,64**	**3,89**	**5 - 24**	**8,71**	**3,16**	**2 - 18**
9 - 11	13	10,39	2,63	5 - 13	10,00	4,20	4 - 17	5,85	2,08	3 - 10
TOTAL	95	18,57	8,21	5 - 45	11,42	3,96	4 - 24	8,31	3,18	2 - 18

NOTE: The 9-11 years of education group was beyond the education range established for this research. However, this information is referred to for comparison and is therefore included in this table.

9.4.2 The effects of age on the performance time of the expanded BGT

In order to consider for possible factors which influenced performance time on the expanded BGT, regression analyses were run to investigate the effect of age on performance time. A regression analysis revealed that no significant correlation exists between age groups and performance time on the expanded BGT for the group as a whole (r=0,004; p=0,97). The individual educational levels also consistently revealed no effects of age on performance time.

9.4.3 The effect of scores on the performance time of the expanded BGT

Multiple regression analysis of variance tests investigated the correlation between test scores and performance time for the expanded BGT, both for the group as a whole and for the individual educational levels. For the group as a whole there was no significant correlation between the performance time and scores (r=-0,04; p=0,66). On the BGT-*c* for the individual educational levels a relatively high correlation existed only for the group with no education at all, which was significant at the 5% level (r=0,59; p=0,02). The correlations for the other educational levels were low and non-significant.

In contrast on the BGRTs, only a low, inverse correlation exists for the group as a whole (r=-0,23; p=0,03). However, when the educational levels are considered separately, no significant correlations exist apart from a moderate inverse correlation, significant at the 10% level for the group with no education (r=-0,48; p=0,068). The correlations for the other educational levels were negative, low and non-significant.

9.4.4 The effects of education on performance time

For each of the BGTs, time steadily decreases as education increases, as can be seen in Table 10 (p.143). A one-way ANOVA revealed that education affected performance time at a 1% level of significance on the expanded BGT (F=12,75, df =4; p<0,0001). Tukey's multiple range analysis tests, presented in Table 11, revealed significant differences between performance time for the five educational levels.

Even where the difference in performance time is not significant between educational levels, these times show a consistent trend within the entire sample. With a single exception, on each of the three tests for each educational level, the performance time is inversely related to the level of education. The exception is for the performance time on the BGRT-*i* where the 0 years of education group takes less time to complete this test than the 1-3 years of education group does. However, the time difference is too small to be of any major significance. As education increases, the performance time decreases, as shown in Figure 3.

Table 11: Tukey's multiple range analysis tests for performance time across education levels on the expanded BGT

Educational level (years)	Educational level (years)				
	0	1 - 3	4 - 6	7 - <9	9 - 11
0		NS	*	**	**
1 - 3	NS		-	**	**
4 - 6	*	-		NS	**
7 - <9	**	-	NS		*
9 - 11	**	**	**	*	

NS = non-significant
- = p ≤ 0,10
* = p ≤ 0,05
** = p ≤ 0,01

Multiple regression analyses investigated the correlation between education and performance time for the expanded BGT and for the individual BGTs. The correlations between education and performance time were high and significant at the 1% level for the expanded BGT and for the individual BGTs as can be seen in Table 12 below. The BGT-*c* revealed the highest correlation between education and performance time. The negative correlation coefficient reveals an inverse relationship between education and performance time. As education increases performance time decreases.

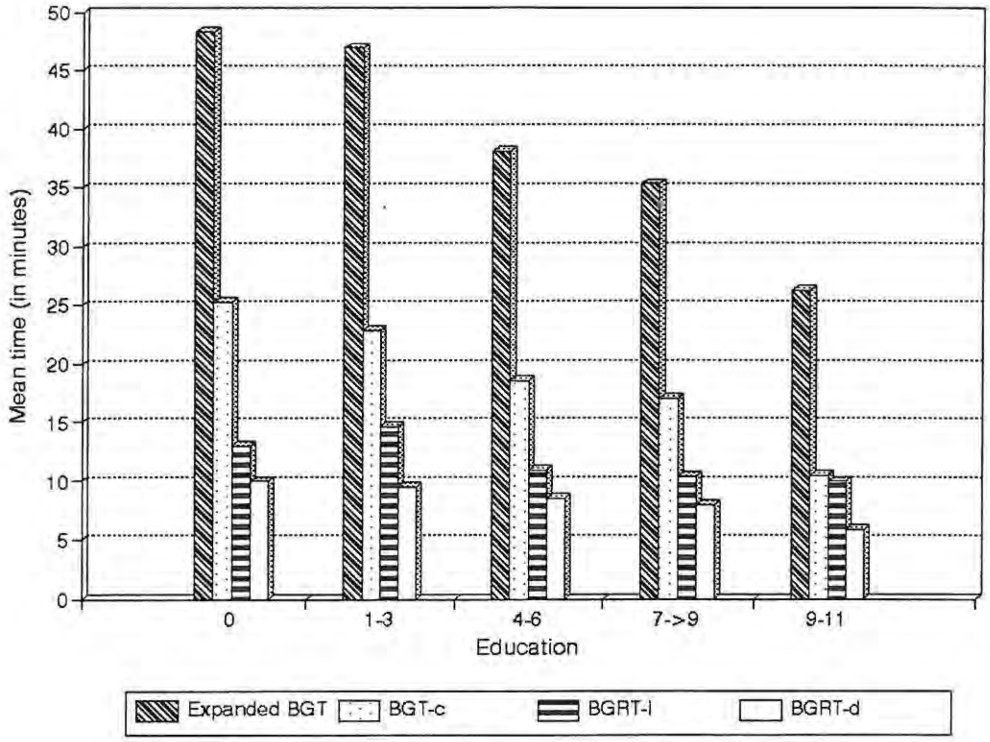


Figure 3: Mean times for the performance on the expanded BGT by educational level

9.4.5 A comparison of performance time on the expanded BGT for non-clinical and possibly brain damaged samples

One-way ANOVA's compared the performance time of the group which has possible brain damage with the group with less than 9 years of education which was used as a normative sample. No significant differences between the groups were found in performance time for the three tests on the expanded BGT. Levene's analysis of variance tests were consistently non-significant. Figure 4 below graphically illustrates the comparison of performance times for the normal and possibly brain damaged samples. On the BGT-*c*, the group with possible brain damage took longer than the non-clinical group, for the groups taken as a whole and for the different educational levels. The one exception to this was the 0 years of education group, where both the possibly impaired and the non clinical group took the same time to complete the BGT-*c*.

Table 12: The correlation between education and performance time for the group as a whole

Expanded BGT			BGT-*c*			BGRT-*i*			BGRT-*d*		
r	*df*	*p*	*r*	*df*	*p*	*r*	*df*	*p*	*r*	*df*	*p*
-0,58	91	0,0001***	-0,54	91	0,0001***	-0,33	91	0,002***	-0,36	91	0,0004***

* $p \leq 0,05$
** $p \leq 0,01$
*** $p \leq 0,001$

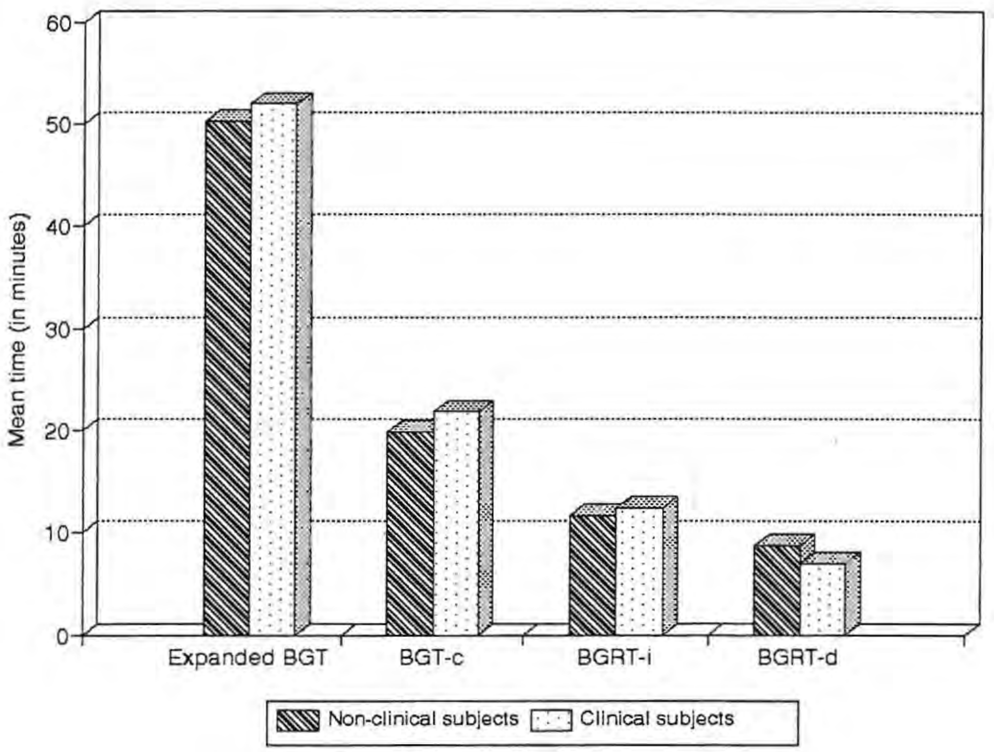


Figure 4: A comparison of performance times of non-clinical subjects with those having possible brain damage

On the BGRT-*i* the possibly brain damaged individuals took less time for two educational levels - those with 0 years of education and those with 1-3 years of education. Mean times for the groups as a whole and the other educational levels revealed that the possibly brain damaged individuals took longer to perform this test.

Time on the BGRT-*d* was shorter for the possibly brain damaged group than the normal group, apart from the 7-<9 years of education group. For the group with 0 years of education on the performance of the BGRT-*d*, the possibly brain damaged group also took less time and this was different from the normal group at a 5% level of significance (*F*=6,91; p=0,0166).

9.5 Comparisons with other normative data

9.5.1 A comparison of the scores on the BGT-*c* with other studies which used Lacks's (1984) scoring system

Multiple *t*-tests compared the BGT-*c* scores in this present study with those in Lacks's (1984) study and Dyall's (1993) study. The results appear in Table 13 below. These were the only studies which could be found which investigated educational levels similar to those used in this study.

Table 13: *t*-test comparisons of BGT-*c* scores in other studies with those in the current study

Study	Educational level (years)	Educational mean (years)	BGT-*c* mean (*SD*)	*df*	*t*-value
Lacks (1984) *versus* current study	3 - <9 0 - <9	NP 4,84	-3,20 (1,51) -4,06 (1,89	155	3,14***
Lacks (1984) *versus* current study	9 - 11 9 - 11	NP 10,00	-2,27 (1,25) -2,77 (1,36)	89	1,36[NS]
Dyall (1993) *versus* current study	2,5 3	2,5 3	 -5,43 (1,72)	42	1,78*
Dyall (1993) *versus* current study	5,5 6	5,5 6	 -4,89 (1,83)	45	3,65***
Dyall (1993) *versus* current study	8,5 9	8,5 9	 -3,08 (1,38)	49	3,49***

NP = data not provided
NS = not significant
* = p ≤ 0,05
** = p ≤ 0,01
*** = p ≤ 0,001

The mean scores obtained in this study for subjects with 0-<9 years of education on the BGT-*c* are significantly different to those obtained by Lacks (1984) for adults with 3-<9 years of education. In contrast, for the 9-11 years of education groups in both studies, there is no significant difference between the BGT-*c* scores. The comparison of the data in the current study with that in Dyall's study (1993) with children of similar educational levels revealed significant differences between the two samples, with the children scoring higher than the adults. Normative data on the BGT-*c* using Lacks's (1984) scoring system for samples with similar educational levels to those in the study in question appear in Appendix 8 (p.262).

9.5.2 A comparison of the BGRTs scores in other studies using identical or similar scoring systems to those used in the present study

9.5.2.1 The BGRTs

The scores of Weiss's (1970) BGRT-*i* and Dyall's (1993) BGRT-*i* and BGRT-*d* for children with educational levels similar to those of the adults in this study, appear in Table 14 below.

Table 14: BGRTs scores from studies for ≤9 years education

Study	Educational level	*n*	BGT-*i*			BGT-*d*		
	Years		*M*	*SD*	Range	*M*	*SD*	Range
Weiss (1970)	± 3	74	1,99	NP	0-5,5			
Dyall (1993)	2,5	37	2,62	1,44	0-6	1,81	1,20	0-4
current study	3	7	1,86	1,07	0-3	1,57	1,40	0-4
Weiss (1970)	± 5	147	2,59	NP	0-5,3			
Dyall (1993)	5½	38	5,39	1,87	1-9	4,97	2,05	1-9
current study	5	9	2,11	1,45	0-5	2,00	1,83	0-6
Weiss (1970)	± 7	90	3,48	NP	0-8,0			
Weiss (1970)	± 9	79	5,63	NP	1,5-8,0			
Dyall (1993)	8½	39	6,39	1,97	2-9	5,68	1,97	0-9
current study	9	12	4,08	2,64	0,9	3,92	2,43	0-8

NP = data not provided

While the mean scores of Weiss's (1970) study and the current study are similar for equivalent education levels, the mean scores of the current study differ from the mean scores of children with similar educational levels as reported by Dyall (1993). Since Weiss (1970) does not provide standard deviations for his data, a statistical comparison of the data in this study with Dyall's (1993) scores of those of Weiss (1970) is not possible. In addition, there appears to be a statistical error in Weiss's (1970, p.144) data as he cites ranges which are not integers, e.g., 5,3. Yet, he made no references to awarding partial scores to any design. It would therefore be expected that only integers would appear in Weiss's (1970) range of scores. Presumably, an anomaly in this data or the reporting thereof exists. A series of *t*-tests were used to compare the mean scores of the BGRTs of adults in this study with Dyall's (1993) scores of children having similar education levels. The results appear in Table 15 below. Apart from the lowest educational group, the mean scores of adults in this study differ significantly from those of children with similar educational levels in the Dyall (1993) study on both the BGRT-*i* and the BGRT-*d*.

9.6 The sample of older adults

Of the five older adults tested, with a mean age of 62,8, only one (20% of the sample) scored within the normal range on the BGT-*c*. The BGT-*c* mean for the group was -5,8 (*SD* =1,79). On the BGRT-*i*, three (60%) of the subjects recalled no designs, one (20%) recalled one design and only one person (20%) scored more than one. The BGRT-*i* mean was 1 (*SD* =1,73). On the BGRT-*d*, four subjects (80%) recalled no designs and one person (20%) recalled one design. The BGT-*d* mean was 0,2 (*SD* =0,45). Regarding time, only one subject finished the BGT-*c* in less than 15 minutes (8 minutes) and performance time ranged from 8 to 50 minutes (*M* =28,2; *SD* =18,89). On the BGRT, the mean performance time was 13,2 minutes (*SD* =5,22) and on the BGRT-*d* it was 6,2 minutes (*SD* =3,96). Overall time for the expanded BGT was 57,6 minutes (*SD* =18,82), with a range of 38-81 minutes. This group had an educational mean of 1 year (*SD* =1,41; range = 0-3 years).

Table 15: *t*-test values revealing differences on the BGRTs between Dyall's (1993) study of children and the current study of adults with similar educational levels

Study	Education mean (years)	BGRT-*i*		*df*	*t*-value
		M	*SD*		
Dyall (1993)	2,5	2,62	1,44	42	$1{,}33^{NS}$
versus					
Current Study	3,0	1,86	1,07		
Dyall (1993)	5,5	5,39	1,87	45	6,59***
versus					
Current Study	5,0	2,11	1,45		
Dyall (1993)	8,5	6,39	1,97	49	3,27***
versus					
Current Study	9,0	4,08	2,64		

Study	Education mean (years)	BGRT-*d*		*df*	*t*-value
		M	*SD*		
Dyall (1993)	2,5	1,81	1,20	42	$0{,}47^{NS}$
versus					
Current Study	3,0	1,57	1,40		
Dyall (1993)	5,5	4,79	2,05	45	3,99***
versus					
Current Study	5,0	2,00	1,83		
Dyall (1993)	8,5	5,68	1,97	49	2,57**
versus					
Current Study	9,0	3,92	2,43		

NS = non significant
** = $p \leq 0{,}01$
*** = $p \leq 0{,}001$

CHAPTER 10: DISCUSSION

This chapter is presented in four sections. The first considers factors relevant to the gathering and processing of data. The next section presents a discussion of the statistical results obtained from this data, which were presented in Chapter 9. The third section takes a critical retrospective look at this research. This chapter ends with a computational formula designed to allocate equal weights to the BGT-*c* and BGRTs on the expanded BGT, in order for a meaningful comparison to be made between the two functions being measured, namely visuomotor perceptual organization and visual memory.

10.1 Noteworthy factors emerging during the gathering and processing of data

10.1.1 Hierarchical process necessary for the recruiting of subjects

Obtaining subjects for this study involved many conflicting tensions and time-consuming negotiation. Eventually, testing took place in each organization through a triangular process of obtaining permission from the employer, the endorsement of each union, and the voluntary consent of each subject. The employer's consent was necessary for two reasons. Firstly, in order to control for brain injury, it was decided that only adults who had been in permanent employment for at least 6 months would be tested (see 6.7: p.112). Secondly, many subjects were not prepared to be tested after work for reasons not central to this study, and testing was therefore mostly conducted during working hours. The need for various unions' approval became clear when prospective subjects stated that they would only consider participating if the union endorsed the project. In turn, union leaders needed the consent of the shop stewards to whom the purpose of testing was explained. Subjects were finally recruited through the shop stewards, with this author meeting with workers where requested, to explain the purpose of testing.

Problems which had to be overcome in dealing with the shop stewards included the following:

(i) The need for assurance that results were confidential, especially regarding employers.

(ii) The need for assurance that although this was a test which identified brain damage, no jobs would be on the line if participants did not fare well. The process of establishing trust between the tester and the shop stewards was lengthy and included attending a number of union meetings;

(iii) The practical problem of time was highlighted. Where testing could not be conducted during office hours, people who were battling to survive financially, dependent on taxis and working long hours, had to be motivated to give up their time;

(iv) It was important and difficult to communicate to the shop stewards that, while endorsement of the project was important, participation should not be made mandatory. This will be discussed further (see 10.1.8: p.163).

Problems and details noted in the recruiting of subjects in this manner, which are relevant for future research include:

(i) The triangular combination of employer, union and subject consent made procuring subjects extremely difficult as this researcher's communication with one organization invariably created suspicion and/or superstition with the other organization and sometimes with the participants themselves.

(ii) The process is a lengthy one. Time was needed for the researcher to establish trust with the union and bridge suspicions because she was white,[1] In addition, time was also required for the union to consider the matter and obtain a mandate from workers before decisions could be made.

(iii) The hierarchial structure of both the employer and the union meant that recruiting subjects and organizing testing times was extremely time-consuming and involved intricate co-ordination and planning. Thus, sufficient time should be allocated for this.

(iv) A pertinent point was that a more casual dress code (similar to that of the workers), as opposed to a more formal dress code (similar to that of the employers), and an informal manner of requesting participation facilitated subjects volunteering.

(v) It is suggested that the company is approached first and not the union. This is illustrated by one instance in which a senior union member obtained volunteers from a company for which he did not work. When this company was approached and union consent was mentioned, they were upset and on these grounds refused permission. This was delicate to convey to the union leader, who by this time was very

1. The union leaders revealed this fact after their trust had been obtained.

enthusiastic about the research. It took much persuasion by this researcher to convince him that a strike was not necessary to demonstrate the union's objection to this refusal.

(vi) A knowledge of union negotiation procedures is helpful as is an understanding of and sensitivity to the sometimes volatile relationships between unions and employers. Researchers should avoid becoming involved in or manipulated by conflict situations between unions and employers.

10.1.2 Problems specifically related to the South African context

10.1.2.1 Racism

Despite the fact that South Africa was dismantling Apartheid when this study began, a number of racial issues emerged. Union officials told the researcher that a mistrust of whites still existed and that the author would have to overcome this to obtain subject participation. A number of entire groups refused requests to participate and cited this mistrust as the reason.

During testing, some incidents occurred which endorsed the hostility and mistrust of some subjects. The most extreme situation occurred on a farm where workers had volunteered to be tested after hours. Despite assurances from the farmer about sufficient chairs, tables and privacy, when this author arrived, a single table and no chairs were provided outdoors. The farmer insisted on being present and when subjects were asked if they smoked more than three reefers of cannabis a day, he threatened that anyone who did would lose their job. Once the BGT-*c* was completed, with subjects standing next to the table and uncomfortably trying to copy the designs, testing was terminated. Illustrative of the more typical type of racism encountered during testing, was an instance where a senior official, who had provided his office for testing, requested that no one sit on his chair, although there was one chair short.

10.1.2.2 Language

South Africa has a broad diversity of languages but fortunately, Grahamstown, as a rural city, does not have large influxes of people from other areas and therefore a lesser diversity of languages. Xhosa was the predominant language, although some subjects only understood Afrikaans. Testing, controlling for normality and particularly briefing were complicated by the fact that there did not appear to be words in the Xhosa language equivalent to words such as 'psychology', 'normative data' and 'meningitis' and other technical terms. Therefore the use of such words was rendered ineffective. This is clearly illustrated by the fact that during the course of testing, three subjects were found to have epilepsy but none knew the name of their disease and referred to it as "fits". Although this language issue is of significance, it is inextricably connected to and further complicated by comprehension difficulties. It will therefore be more fully discussed with issues of vocabulary and comprehension in section 10.1.3.2 (p.156).

An issue on the semantics of language also emerged. On the commencement of the BGRT-*i*, the tester's instruction to subjects was to "draw the designs which you first drew from memory". Once translated, in almost all cases, this instruction was perceived by subjects as "draw the first design again". The instructions had to be changed and subjects were instead asked to once more draw the designs which they had already drawn.

In summary, researchers in South Africa should be aware that a mistrust of whites exists and they should also be careful not to align themselves with racism in the course of conducting research. Another specifically South African factor to note is the differences between English (or Afrikaans) and African languages which highlight the need for sensitivity to language subtleties.

10.1.3 Problems relating to low-educated subjects

Testing low-educated adults involved a number of problems. These included issues surrounding superstitions, vocabulary and comprehension, answers to questions regarding the normality of subjects, subjects tiring, the drawing task, education-related problems and the need for debriefing.

10.1.3.1 Superstitions and suspicions

Superstitions were rife among the population which was being studied and overcoming these was sometimes difficult. In some cases superstitions could not be overcome and recruiting of certain groups was abandoned. Superstitions about psychology were particularly rife, and many subjects expressed fears about the researcher being able to 'read minds' and 'cast spells' on subjects who participated. Some believed that participation would cause enduring emotional or physical problems later. In some cases superstitions appeared to be connected to subjects' experiences of being victims of racism, which elicited suspicion of white people. Initial suspicion was also aroused because this author communicated with the employer. This was aggravated by the fact that at the time of testing, fundamental changes in some organizations' structures were anticipated due to changes in the country's political structure. A number of workers refused to participate as they believed that the author was working for the employer and that those who participated in the study would be retrenched as they allied themselves with 'whites'. Other more obscure beliefs expressed by prospective subjects included the belief that participation would endanger the security of their positions in an unknown way.

10.1.3.2 Vocabulary and comprehension problems

A problem specifically related to the low-educated sample group centred around vocabulary and comprehension. This made both briefing the subjects as to the purpose of this study, and asking questions which controlled for normality, difficult. The briefing, which was initially given to the shop stewards, as laid out in point form below, was explained to each participant along the following lines:

(i) the importance of differentiation between organic and functional impairment for treatment and outcome;

(ii) the concept of test norms and the fact that although the BGT identifies brain damage, this did not mean that brain damage was the criterion being sought in the subjects;

(iii) the aesthetic value to participants, that while they would not benefit in any way by doing the tests, others would;

(iv) the concept that nothing would happen to the individuals being tested during or after their participation. (This is where superstitions were rife and it was very difficult to explain in a manner which catered for the viewpoint of a low-educated person. Subjects were first told that drawing 'a few patterns' on a piece of paper could identify brain damage, i.e., an invisible condition within the head. This concept in itself appeared to be 'magical' to the least educated subjects. Then, in apparent contradiction, they were told that drawing the designs would not harm them, nor enable the tester to 'see' the thoughts in their head.)

The language issue, as already discussed (see 10.1.2.2: p.154), further compounded these problems. The problem of language and comprehension was overcome to an extent by (i) altering the terminology used to that with which subjects were familiar, and (ii) developing a fairly lengthy descriptive and illustrative method of briefing and asking questions to control for normality. For example, the concept of normative data was explained by making analogies to the work done by those particular subjects. A group of electrical assistants was told that they were able to identify a faulty or broken cable as opposed to a 'good' cable when searching for an electrical fault. It was explained that they, as subjects, were helping the tester to establish what exactly such a 'good' cable looks like. While this was not entirely satisfactory, no other option was available. This author was aware that this difficulty in controlling for possible neurological damage could mean that some individuals with neurological impairment might be included in the sample. However, no other approach was available and at least this method enabled 33 subjects to be excluded from the normative data. The briefing time was doubled by the need for translation.

10.1.3.3 Problems with the responses to questions used to control for normality

A problem which arose with the questions which controlled for normality was what appeared to be either a willingness to please the tester or, alternatively, a matter of pride on the subjects' behalf. This pertained specifically to stab wounds and subjects were anxious to record these, particularly where a scar was visible. This problem was solved by asking those subjects who reported head injuries due to stab wounds, to elaborate on the incident after the testing. In many cases, it was discovered that the subject had in fact sustained a stab wound, but to a part of the body other than the head.

10.1.3.4 Subjects tiring

The sample in question was clearly not used to sitting for long periods of time. A possible reason for this is that most of these subjects worked in physically active occupations. Further, for those with no education or very low education, the tasks demanded high levels of concentration with which subjects were clearly unfamiliar. Also, as already stated (see

5.1.4: p.94) subjects also were unfamiliar with using pencils and paper. Some subjects, particularly those who took a long time, got restless towards the end. The tester found it helpful to tell subjects that the BGRT-*d* was the last task.

10.1.3.5 The need for debriefing

Some groups expressed feelings of being 'tricked' by the recall since no warning was given. Again, in some instances there were racial undertones to these feelings. Here a debriefing was necessary and subjects were told that this was done to prevent concern during the BGT-*c*, which would affect performance. This explanation proved sufficient.

10.1.4 Problems relating to the location of testing

Ideally, a quiet place is needed for testing (see 5.2.2: p.97). Yet, despite the fact that this author ensured beforehand that there were chairs and tables in a quiet area, these criteria did not always materialize. Some interruptions were brief and legitimate, but others were not, and some had racial undertones. Fortunately in most cases, interruptions occurred during the briefing stage and not during the actual testing. Testing low-educated individuals, many of whom are labourers, at their workplace is not always ideal. Smooth wooden planks formed part of the testing kit and where table surfaces were rough, these were placed on top of the tables. Drawings executed on a rough table surface can simulate Lacks's (1984) error of motor incoordination. However, most of the problems experienced were still preferable to removing subjects from their work environment, which might have increased their anxiety levels. Anxiety, as already discussed (see 5.1.2: p.93), needs to be seriously considered in the sample tested.

10.1.5 Group testing: advantages and problems experienced

10.1.5.1 Advantages of group testing

The group administration of the expanded BGT was valuable in terms of saving time. The test was fully administered to 128 subjects with a mean performance time of 48,3 minutes. Using the individual method, testing would have taken approximately 103 hours. Instead, the application of the group method took about 19,5 hours which was one-fifth of the time which the individual method would have taken. This concurs with Becker and Sabatino

(1971) who reduced testing time by one-third, using the group method. The above excludes the briefing and debriefing times, which have already been discussed (see 10.1.3.2: p.156). If these were considered, the time saved would be substantially more.

10.1.5.2 Group dynamics

It was discussed earlier (see 5.1.4: p.94) that school children being tested revealed strong group dynamics. No such effects were evident in the groups of adults tested in this research. The most likely reasons for this is that the time taken to complete the protocols is correlated to educational levels and this is discussed later (see 10.2.2: p.166). While the groups of school children tested were of the same educational standard, most groups in this study consisted of subjects with a broad range of educational levels.

10.1.5.3 Problems with group testing

Despite the advantages of group testing, certain problems were experienced. A number of factors need to be considered when using this method of administration.

Firstly, testing low-educated adults is very different to testing school children. School children are familiar with a teacher's authority, adhering to instructions, and being part of a group. For adults, this was not true. Gathering together in a new situation was often seen by the subjects in this study as a time for socializing. Also it is conjectured by this author that anxiety played a role as excitement levels were sometimes high. Without becoming authoritarian, group chatter was a problem. Limiting groups to a maximum of 14 subjects, and avoiding certain times for testing, helped to control this problem. For example, after discussing these problems with union leaders, they advised that payday, the day following payday and Friday afternoons were not recommended for testing. Testing of one group on the morning after payday was stopped when it was realised that a number of subjects were leaving the room to drink alcohol.

Secondly, this researcher found that placing the BG designs in front of subjects prior to the briefing and the answering of questions was problematic. Subjects often looked through them, changed the order and orientation, or began copying them. Cards then had to be reorientated. This problem was resolved by only giving the cards to subjects as testing began.

Thirdly, timing and observation in a group context is highly challenging, particularly since the BGRT-*d* commences precisely ten minutes after the completion of the BGRT-*i*. It is also very easy, in a group where individuals are working at their own pace, to lose track of the stage of testing at which each subject is, and which test must be administered next. The numbering and coding system developed for group testing and used by this researcher (see 8.3: p.130 and Appendix 6A: p.260) eliminated timing and observational errors.

Fourthly, dealing with anxiety in a group situation also posed a problem. When intense anxiety occurred, which was normally on the BGRT-*i* and occasionally on the BGT-*c*, immediate termination of testing in the group context would have singled out the individual. In most cases, these subjects were left to draw a man.

Finally, since space was sometimes a problem, screens were used to prevent subjects seeing others' designs on the BGRT-*d*. Early in the testing process it was clear that some subjects became anxious and/or suspicious when the researcher erected the screens. This was solved by asking subjects to erect the screens themselves so they could not see the paper of the other subjects in the group. The feeling of control this provided probably assisted to eliminate the extant anxiety.

10.1.5.4 Guidelines for the group testing of low-educated adults

In summary, the following guidelines are suggested to facilitate group testing with low-educated adults:

(i) When the administrative procedures used in this thesis are adhered to, an *experienced* tester could handle a maximum of eight subjects alone, or 14 when a research assistant is present. Larger groups are not advisable. Initially, smaller groups are suggested;

(ii) Testing should be conducted at appropriate times in order to avoid distractions;

(iii) Sufficient time should be allocated between groups to reorientate the cards;

(iv) If possible, group members should have similar educational levels to obtain some sort of parity in the testing duration;

(v) Where screens are used, subjects should be asked to erect these themselves;

(vi) Each person should be given a number and protocols should be both numbered and coded as each test is begun;

(vii) Subject sheets should be used for observational and timing accuracy;

(viii) Cards should be placed in front of individuals just prior to the start of testing.

10.1.6 Encouragement

As with the previous study, the author of this paper found subjects were quite adamant that they could not remember any more designs or that they required more time as they thought that they could remember more. Thus encouragement as used by some researchers (Rogers *et al.*, 1975; Hutton, 1966) appears to be unnecessary. It is suggested that for future research, no encouragement is used, as standardization of this procedure is difficult.

10.1.7 Scoring

10.1.7.1 Time taken to score the expanded BGT

Lacks (1984) warns that the BGT can appear deceptively simple and the same warning can be applied to her scoring system. Although it is simple, it has many subtleties that are often only evident once that specific problem appears in a protocol. An error a novice scorer may make is to overlook the fact that each scoring item is not necessarily applicable to each design. The time for scoring was under three minutes per protocol on average, as Lacks (1984) states, provided scoring was done in batches with undisturbed time. The scoring of the two BGRTs took an estimated three minutes altogether. Obviously some protocols were problematic and took longer to score.

10.1.7.2 Problems in scoring rotations on the BGRTs

The three BGT protocols of each person must be kept together in order to score rotations correctly. Where the researcher has marked that a BGT-*c* design was rotated due to the card or paper being rotated (see 5.2.4: p.99), this precise orientation is taken as the original stimulus for the BGRTs. For example, if design A is drawn upside-down because the card was upside-down, a recalled design A would be scored correct only, if it too, was drawn upside-down.

A more difficult problem emerged which raises questions concerning the scoring of recalled rotations when the BG designs and the paper were correctly orientated on the BGT-*c*. For example, if a subject rotates design 2 by 180°, what is correct on the recall protocol? If the subject draws the figure as rotated again, is this correctly recalled, or is it an error? Even more problematic is the rotation of a design on the BGT-*c* which is followed by a correctly orientated design on one of the BGRTs. Two explanations for this are possible: (i) the subject could have corrected the rotation with more time to process the gestalt; or (ii) the subject could have rotated the design again by a further 180°, thus returning the figure unknowingly and unintentionally by a full 360°, to its original orientation.

This problem was not resolved. In this research, the method adopted before the problem was evident, was used i.e. only correctly orientated designs were scored as correct. However, this was not entirely satisfactory. A method needs to be found of establishing whether a second 180° change in the orientation of an already rotated design is due to a correction, or a second rotation. One scoring option would be to delete this item from the BGRT scoring altogether, which would be regrettable. As already discussed, rotations are both diagnostic signs (see 2.1.5: p.12) and their neurological significance on the BGT is well supported by literature (see 7.5.1: 118). A compromise, until further research can resolve this problem, could be to eliminate this scoring item only from the recall of subjects who rotate designs on the BGT-*c*.

10.1.8 Ethics

10.1.8.1 Informed participation

In keeping with the philosophy of this author, ethics played a prominent role in this study. Firstly, "informed participation", a term used by Marteau (1994, p.1), was strictly adhered to. Despite the length of time it took to provide an 'informed' briefing and the refusals that sometimes resulted from such a briefing, all subjects were fully informed of the study. Nor was any subject persuaded or coerced to participate in any way by this researcher. It would have been contrary to ethical standards not to fully inform participants as to the purpose of the testing and irresponsible to test a subject who believed participation would produce some adverse effect at a later stage. Although the researcher knew this belief to be incorrect, such a belief could cause considerable distress to the subject.

Further, the researcher gained the impression that certain union members became overzealous in procuring subjects and were possibly pressurizing members to participate. For these reasons, no testing began before all subjects were offered the opportunity to withdraw, with the assurance there would be no negative consequences. Although only one person utilized the opportunity to withdraw, this indicated the necessity of providing this option. On the other hand, it was also made clear that their participation would not provide any positive benefits particularly regarding their occupational status.

10.1.8.2 Commitment to the ethos of psychology

Although no promises of benefits were made, this researcher did undertake to try and assist or make referrals to assist subjects with problems, where this was within her ability. This offer was motivated by this author's perception that research does not only involve a commitment by subjects to the researcher, but also a commitment by the researcher to the subjects. In this regard, Fisher (1993) stresses that the ethical need of researchers to activate a process of referral for research subjects who require psychological treatment. This is congruent with the ethics of the psychology profession which is the "promotion of human welfare" (Fisher,1993, p.381). She adds that failure to attend to such matters is a failure in the responsibility of researchers to care for the wellbeing of the individual.

In the course of this study, where requested, referrals were made. Three referrals were made to Fort England, (a regional mental health hospital), two of which involved untreated epilepsy. Three subjects were counselled and then given information about the Rhodes Psychology Clinic as offers of a direct referral were turned down. Attention was given to other individual problems. Beyond the limits of psychology theory, but not beyond its ethos, this author assisted about 15 individuals in having their eyes tested, and where necessary, acquiring spectacles at lower fees and gaining the financial support of the employer. Since the lack of medical treatment amongst subjects appeared to be a prime cause of concern, all subjects were notified about the presence of the Health Train when it was in the city, which offers medical care at a very reduced cost. Finally, attendance at literacy courses was facilitated by this researcher where subjects requested such assistance.

10.2 Discussion of statistical results

From the results of these data, it is clear that significant educational effects exist on the expanded BGT, both in terms of scores and the performance time. As hypothesized earlier (see Chapter 1: p.1 and 2.2.2.3: p.36), existing norms on the BGT for well-educated adults cannot be appropriately applied to individuals with low education. The meaningful data obtained in this study also suggests that the use of Lack's (1984) scoring system with uneducated adults is appropriate. To the author's knowledge this has not been done before.

The discussion below first addresses the demographic variables of age and gender, in order to exclude these as possible influencing factors. Next, the educational effects on the expanded BGT scores are considered and Lacks's (1984) neurological 'cut-off' point is discussed, in relation to this study. Then, the high correlations between the three tests are discussed followed by a consideration of the performance time on the expanded BGT. The mean scores obtained in this study are then compared with other studies and the similarities and differences are discussed. The scores and performance times for older adults are then considered. Finally, the need for normative data for low-educated adults is discussed.

10.2.1 Demographic data

10.2.1.1 The effect of age

Lacks (1984) states that subjects aged 17-59 consistently score above the neurological cut-off point of -5. The trend in Lacks's (1984) data is that the highest scores are obtained in most cases by the 25-34 year olds. The 17-24 year olds score less than this group, but the 55-59 year old group has the lowest scores overall. Thus, apart from the 17-24 year old group, Lacks's (1984) younger subjects perform better than older subjects. The mean ages of subjects in the present sample fall approximately midway on Lacks's (1984) age range. Thus extreme scores attributable to age are not likely to be present in the current sample. Lacks (1984) categorized her age groups in 10-year spans (e.g. 35-44 years old), except for the youngest and oldest groups, which are categorized according to five year spans.

The results of this study revealed that significant age differences exist between some of the educational levels (see 9.1.1: p.135). However, the range of the mean ages for each group is only 11,6 years (minimum value = 34,2 years; maximum value = 45,8 years). In light of the above, the range of the means of the five educational groups is sufficiently low for the variable of age not to exert a major influence on these findings.

10.2.1.2 The effect of gender

Although not a primary objective of this study, it was decided to investigate the influence of gender as no such published research for low-educated adults on the BGT could be found. No gender effects on the BGT-*c* were found. This contrasts with some neuropsychological tests, especially motor tests, which revealed gender effects (see 6.2: p.105). However, the lack of gender effects are consistent with most research findings which failed to find gender effects on the BGT in more educated adults (see 6.2: p.105). The lack of gender effects suggests that the research conducted on more educated adults is, in respect of gender differences, applicable to low-educated adults. There was also a lack of gender differences between the scores of both BGRTs. Further, no gender effects were revealed for performance time on the expanded BGT. As already discussed (see 6.2: p.105), few memory tests reveal gender effects on test scores. The investigation into gender effects in this study is useful, because, as already stated, no published data concerning the performance of low-educated adults on the BGT exist, apart from Lacks's 3-<9 years of education group. Considering the above, the decision (see 6.2: p.105) not to control for gender in this study was appropriate.

10.2.2 Education effects on the expanded BGT scores

10.2.2.1 The BGT-*c*

Scores improve as education increases. Consistent with research cited (see 2.2.2.2: p.34), the findings in this study clearly reveal an educational effect for the BGT-*c*. Apart from the anomaly of the scores of the 0 years of education group (which is discussed below), the scores improve as education does (see Table 4: p.138). Significant education effects can be seen in Table 6 (p.139). These results are endorsed by the significant differences in scores between those with 0-<9 years of education and those with 9-11 years of education (see Table 6: p.139). The greatest rate of change of scores between groups occurs between subjects with 4-6 years of education and those with 7-<9 years of education. Above 7 years of education, scores even out into a plateau with no significant differences between the scores of the 7-<9 and 9-11 year educational groups.

In this study the 0 years of education group scored especially well on the BGT-*c* and no significant differences exist between these scores and those of the 7-<9 years and 9-11 years of education groups. This contrasts with the study by Grossi *et al.* (1993). As already discussed (see 2.2.2.2: p.34), this study found that individuals with no more than 3 years of education performed significantly better that those without any schooling. However, a difference between this study and that of Grossi *et al.* (1993) is that the latter study investigated educational effects in a sample of older adults. Older adults perform less well on neuropsychological tests as neurological deterioration due to aging begins in some (e.g. Hart, 1993; Lacks & Storandt, 1982. See also 6.4: p.108).

A few possible reasons exist for the inconsistency of scores for the group without any education. Firstly, as already discussed (see 8.1: p.127), in order to control for normality, a subject selection criterion was that all subjects had to have been employed for at least 6 months. Thus, the illiterate subjects in this study were all employed despite the high unemployment rate in Grahamstown. This author suggests that individuals without education who managed to be appointed to and retain an occupational position for more than 6 months, are likely to represent the top echelon of such a population rather than being representative of a normal population distribution. Based on this supposition, the IQ range of the subjects without education in this study is likely to be at least above average. As already discussed (see 6.3: p.107), correlations between BGT performance and IQ in educated adults using Lacks's (1984) system of r=-0,33 and r=-0,13 have been found (Brilliant & Gynther, 1963; Lacks, 1984) and of r=-0,24 in children on the expanded BGT with a non-significant correlation of p=0,35 (Dyall, 1993). Lacks (1984) does not mention the probability coefficients related to her correlations. No data regarding correlations between IQ and BGT performance for uneducated individuals could be found in published literature. However, in light of the above, the high scores of the uneducated sample in this study might be slightly, but not solely, attributable to the effect of higher than average IQs.

A second, more plausible explanation, related to the above, could be that the uneducated sample in this study consisted of individuals who are highly motivated. Those participating in the study who had no education at all might have been very intent on proving their abilities on the expanded BGT to compensate for their lack of literacy skills. This probability is supported by the fact that many of those without education expressed concern at not being able to write and continually asked whether the test would still be applicable to them. Lacks (1984), as already discussed (see 5.1.4: p.94), states that a neurologically impaired individual may score within the normal range if extreme care and excessive time is taken. In other words, extreme care and time can improve BGT-*c* scores. The 0 years of education group in this study, took the longest time to complete the BGT-*c* which could imply a special amount of care taken by this group. In addition, this group is the only one which revealed a relatively high correlation between scores and performance time, which further supports this supposition (see 9.4.3: p.144).

A third explanation is that this author failed to establish whether subjects had attended or were attending literacy classes. This is discussed later in this section (see 10.3.1: p.194). In other words, despite a lack of formal schooling, these subjects may have received educational development through adult education or career-related training provided by employers. Considering the educational effects on the expanded BGT, subjects who have a degree of literacy would obviously perform better. However, this is unlikely to explain the total effect as certainly a number of subjects required assistance in writing their names, so clearly at least part of the sample had not attended literacy classes. Also, if this explanation sufficed, then the scores obtained by this group might be expected to be equivalent to the educational group with 1-3 years of education, however, they are not.

Fourthly, a further factor which could account for these findings is that the lack of any education at all is likely to be due to the absence of opportunity, albeit geographical or financial. By contrast, those who dropped out of school within the first three years and have 1-3 years of education, obviously had the initial opportunity to attend school. The subjects with 1-3 years of education may have discontinued their education through a change in circumstances relating to opportunity. However, there could be other reasons for their discontinuation of schooling. Such individuals might not have been able to meet the standard required of them, due to a low IQ or an undetected retardation. It is possible that many who left school at this stage discontinued their education through a lack of ability rather than a lack of opportunity. This would mean that while the sample with no education probably represents a normal distribution in terms of ability, those with 1-3 years of education are more likely to be over-represented, to an extent at least, by low IQ or mild retardation. As already stated (see 2.1.5: p.12), neurological problems in children often first become evident when schooling begins. A factor which supports this supposition is the discrepancy in the ranges of the mean scores between the two groups. For the subjects with no schooling this was -1 - -7, while scores of the 1-3 years of education group ranged from -4 to -9. The highest scores for subjects with 1-3 years of education was substantially poorer than the highest scores of those with no education. Further, the mean score of the entire 1-3 years of education group was -5.53 (see Table 6: p.139) which fell within Lacks's (1984) range for neurological impairment as her cut off score is -5. The mean score for the 0 years of education group was significantly higher at -3,73 (see Table 6: p.139).

10.2.2.2 The BGRTs

Across the educational levels for both BGRTs, the scores increase as the educational level does (see Table 4: p.138). Again, the exception is the 0 years of education group which still scores better that the 1-3 years and 4-6 years of education groups, but on the BGRT-*i*, unlike the BGT-*c*, this score mean is significantly lower than that of the 7-<9 years of education group. Possible reasons for this anomaly are the same as those suggested for the anomaly on the BGT-*c* (see 10.2.2.1: p.166), namely a higher than average IQ, a high level of motivation and extreme care. Interestingly, Schneider and Pressley (1988) suggest that motivation could influence rehearsal, which improves memory abilities. Although no published research could be found which investigates the educational effects on the BGRT(s), the influence of education on memory is consistent with research cited regarding neuropsychological memory tests in general (see 2.2.2.2: p.34). In this section, Lezak (1983) stated that research has established that education is positively related to performance on neuropsychological tests even where functions are apparently unrelated to academic performance, such as with immediate memory tests.

Schneider and Pressley (1988,p.48) conclude from a review of research that the process of memory differs between younger and older children in a "qualitative" rather than a quantitative manner. They add that processes indicative of rehearsal strategies are evident in children from about 8 to 10 years of age. They state that "all of the relevant data suggest that the developmental increases in active, cumulative rehearsal play a crucial role in the explanation of age differences in free .... recall tests" (Schneider & Pressley, 1988, p.50). They add that older children have "more concepts and interconceptual associations to trigger [the] use of strategies" (Schneider & Pressley, 1988, p.51). On the other hand, younger children have difficulty because of "the mental effort to carry out cumulative rehearsal [and] inefficient use of strategy components" (Schneider & Pressley, 1988, p.51). It is possible that this suggests an education rather than an age effect since education and age in school-going children is highly correlated (see 6.4: p.108).

Coinciding with scores on the BGT-*c*, the mean scores on both BGRTs show a sharp increase in the rate of change between the 4-6 years of education and the 7-<9 years of education groups. Again, similar to the BGT-*c*, the scores appear to level out at higher educational levels with no significant differences between the BGRT scores of either of the two upper educational levels. Thus, as education increases, the disparity between education levels for the BGRT-*i* and the BGRT-*d* scores decreases, again creating a plateau for the groups with more than 6 years of education.

Congruent with other research, scores on the BGRT-*d* were less than those on the BGT-*i*. This is in keeping with the original rationale of the test as well as with more recent research on neuropsychological tests. Bender (1938) stated that with the passing of time, gestalten revert to more primitive forms. Lezak (1983) observes that errors in a copy or an immediate recall test may be exaggerated in a delayed recall test (see 2.3.6.2: p.43). Wilkinson and Koestler (1983), as already discussed (see 2.3.6.2: p.43), state that the chance of a design being recalled increases if it has been previously recalled and decreases on successive recalls if it has not been previously recalled.

Noteworthy is that the scores for the group with 0 years of education, which are unexpectedly higher than those for groups with 1-3 and 4-6 years education for both the BGT-*c* and the BGRT-*i*, are equalized on the delayed version of the expanded BGT. On the BGRT-*d*, the 0 years of education group scores less than those in the 1-3 years of education group. Literature reviewed supports this result to an extent. As already discussed (see 2.3.6.2: p.43) a delayed memory test can be more sensitive to neurological functioning than an immediate memory measure. It is possible that a delayed memory test after an interference procedure is testing a factor which is not as easily influenced by those factors which operate on the first two versions. It was discussed above (see 10.2.2.1: p.166) that a possibly higher than average IQ, motivation and extreme care partly accounted for the results of this group on the BGT-*c* and the BGRT-*i*.

This equalization of scores on the BGRT-*d* supports the possibility that it is measuring a different facet of memory which appears less vulnerable to extraneous influences suggested to be IQ, performance time, effort and motivation (see 10.2.2.1: p.166). This finding also supports the diffuse nature of memory and concurs with literature which states that memory is not a unitary function (see 2.3.2: p.38). In this regard, Kampen and Grafman (1989, p.57) endorse the multi-functioning nature of memory and add that different memory tests measure "different types of memory in different domains". Sabatino and Becker (1969) correctly note that since the designs are the same for the BGT-*c* and the BGRT, the differences in scores can be attributed to the different functions which are being measured. If this concept is applied to the two BGRTs, the significant difference in the scores between the two tests (see 9.3: p.142) appears indicative that to some extent, different aspects of memory are being measured. From these findings, it appears that the BGRT-*d* is specifically sensitive to a classroom effect, as opposed to other kinds of learning and intellectual development. In a school class, children are almost continually required to rehearse knowledge for recall, be it mathematics, spelling or poetry. According to some memory models, rehearsal strengthens memory traces (Baddeley, 1987). It is suggested that perhaps, with school learning experience, individuals unconsciously rehearse the BG designs while they are drawing them, but that individuals with no education and therefore no classroom experience do not. This would account for the differences found in the performance of those with no education on the BGRT-*i* and the BGRT-*d*. It could also be possible that while motivation might have increased rehearsal and therefore memory, on the BGRT-*i*, the ineffective use of rehearsal strategies prevented this retention from being sustained over a delayed time period.

10.2.2.3 The comparison of the scores on the expanded BGT of non-clinical subjects with those having possible brain damage

The significant difference between the BGT-*c* scores of the non-clinical subjects and those with possible brain damage (see 9.2.2: p.139) provides an indication that the BGT-*c*, scored according to Lacks's (1984) system, may well be appropriate to low-educated adults as suggested earlier (see 6.5: p.110). However, as already stated, the sample numbers are disproportionate and further research is needed to verify the implications of this data. The lack of significant differences between these two groups for the BGRTs must be seen in the same context as the above data, namely that sample numbers are small and disproportionate. The fact that scores of the BGRTs are consistently lower for the possibly brain damaged group indicated a possible sensitivity of the BGRTs to neurological functioning. Certainly, the potential of the BGRTs need to be explored further.

10.2.2.4 The consideration of an appropriate neurological cut-off point and base rates for low-educated adults

(i) The consideration of a neurological cut-off point

Considering that normative data for those with low education is needed (see 10.2.2: p.166), it is logical that existing cut-off scores on the BGT-*c* for educated adults cannot be appropriately applied to low-educated adults.

If Lacks's (1984) cut-off score of -5 were applied to this sample, 34 or 41,46% of the non-clinical subjects would be misclassified as neurologically impaired and 71,4% of those with possible brain damage would fall into the neurological range (see Table 7: p.141). Less maturity and experience in individuals influences the gestalt and forms part of the original rationale of the test (Bender, 1938). Thus for subjects with low education, it could be expected that the cut-off score for organic impairment would need to be lower than Lacks's (1984) cut-off score for more educated adults. Lacks (1984) stated that 21% of her low-educated sample (3-<9 years of education), scored within her neurological impairment range. It is of interest to note that when 19,5%[2] of the poorest scores were removed from the non-clinical sample, in fact all scores of -6 and higher were removed. Considering the lack of medical care, the high incidence of head injury in the Eastern Cape (see 8.1: p.127), and the problem with controlling for normality (see 10.1.3.2: p.156 and 10.3: p.194), it is possible that about 21% of the subjects might have neurological impairment. It appears from this preliminary information that a neurological cut-off point might reasonably be around -6. Obviously, this is a rough estimate based on small sample numbers and extensive research with larger samples would be necessary before a cut-off score could be applied with any confidence. If a -6 cut-off score were applied to the 'corrected' sample above, none of the non-clinical subjects would fall into the organic range. If a -6 cut-off point is applied to the present sample, 19,58% of the non-clinical sample would fall into the brain damaged range, while 61,9% of those with possible brain damage would. If it is hypothesized that none of the non-clinical sample have undetected neurological impairment and that all those in the possibly brain damaged group do in fact have brain damage, then a 71,16% diagnostic accuracy has been achieved if a -6 cut-off score is applied. In reality, this accuracy is likely to be higher since it is probable that some subjects in the non-clinical group have neurological impairment (see 10.1.3.1: p.155 and 10.3: p.194). Further, it is possible that not all of those who reported possible neurological impairment, are in fact impaired.

Considering the inappropriate cut-off score and the educational effects on the expanded BGT, separate normative data specific to the population in question is needed. This is consistent with research literature which is discussed later (see: 10.2.7: p.190). Further research is now indicated to establish scores on the expanded BGT for neurologically- and functionally-impaired adults with low educational levels. However, in establishing cut-off scores, base rates must also be considered.

---

2. Due to sample size, this is the closest appropriate percentage to Lacks's (1984) figure of 21%.

(ii) Neurological cut-off points and appropriate base rates

Bornstein (1986, p.413) states that neuropsychological tests gain meaning by equating performance of an individual with an established "cut-off score". However, research clearly indicates the importance of also establishing base rates, i.e. the frequency of the conditions being investigated in the specific type of population in which testing is conducted in order for cut-off scores to be useful (Meehl & Rosen, 1955; Rimm, 1963; Heaton *et al.*, 1976). The incidence of head injury reported by subjects in this study (see 8.1: 127) constitutes 22,65%[3] of the subjects who volunteered. However, during the first half of testing subjects, the incidence of possible brain damage in subjects approached 40%. Because of this, this author specifically requested volunteers who had not sustained head injuries when recruiting subjects during the second half of testing. Thus, the actual incidence of head injury could be substantially higher than 22,65%. If the high incidence of head injuries is representative of this population group, which only further research can clarify, then base rates must receive particular consideration in the establishment of the neurological cut-off point for this population.

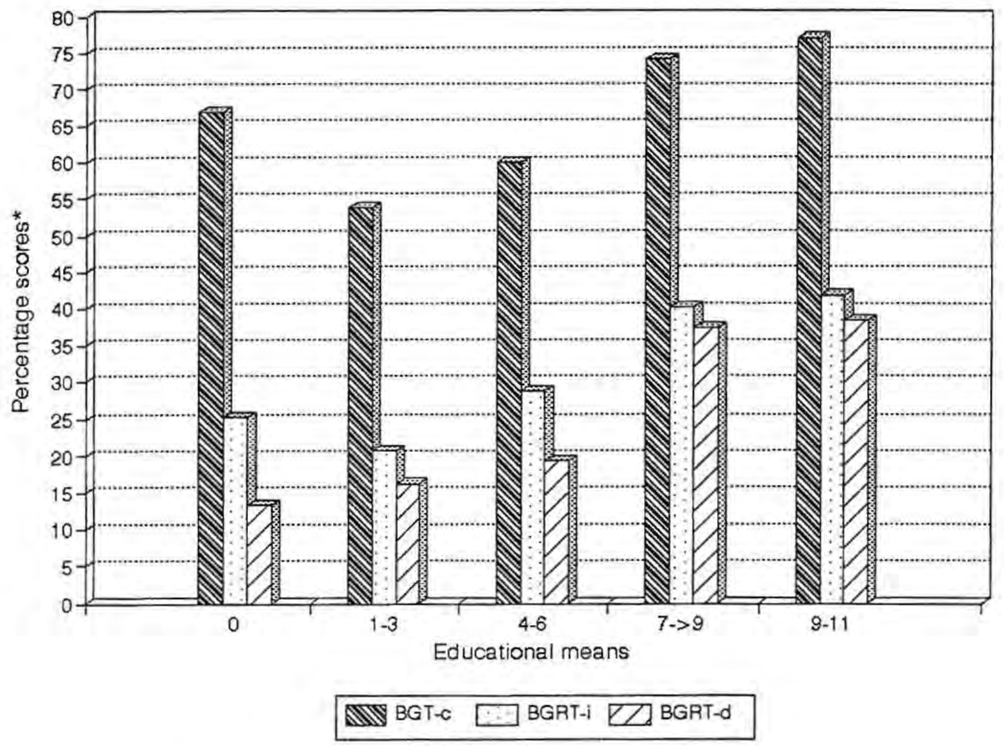
10.2.2.5 The plateau effect

Apart from the anomaly in the scores of the group without education, which has already been discussed (see 10.2.2.1: p.166), there is a sharp, disproportionate rate of increase between the expanded BGT scores of the educational groups with 6 years of education and less, and those with 7 years and more education for all three tests. Above the 6 years of education level the scores even out, and there is little difference between the scores of these two educational levels. Statistical analyses revealed that no significant differences existed between the scores of all the tests on the expanded BGT between the two educational levels which represent more than 6 years of education (see Table 6: p.139). Further, it appears that above 6 years of education, scores reach a plateau. This plateau effect can be clearly seen in Figure 5 below. The increase in the rate of change is unlikely to be due to an age effect since, as already discussed (see 10.2.1: p.165), the range of the

---

3. From informal inquiries into the apparent high incidence in head injuries, it appears that a culture of stick fighting which exists in the Eastern Cape could, at least partially, account for this. Stick fighting is a popular game in which individuals aim to render opponents unconscious by hitting them on their heads with solid wooden sticks.

age means for each group are considered too small to exert a significant influence on test scores.



* The conversion of scores to percentages is discussed in 10.4 (p.196)

Figure 5: The plateau effect

Another consideration is the effect of IQ. In this regard, Lacks (1984) notes that the effect of education on scores only gains clinical significance where education is less than 8 years or where IQ is very low (see 6.3: p.107). The sharp rate of improvement in scores is also unlikely to be an IQ effect since it is highly improbable that the IQs of all those with 6 or less years education are substantially lower overall than all those with 7 years or more education as subjects were recruited in the same manner. Also, it has already been suggested that the group without education might have a higher than average IQ while the 1-3 years of education group could have a lower than average I.Q. (see 10.2.2.1: 166). Yet apart from the anomaly on the BGT-*c*, this increased rate of change is applicable to this group as well. As already stated (see 10.2.2.1: p.166), the correlation between IQ and performance ranges between r=0,13 and r=0,33 which would be insufficient to cause the extent of the rate of change revealed on the expanded BGT scores. Koppitz (1975) supports this argument with her findings that BGT-*c* scores vary between subjects even when IQ is held constant. It is clear, therefore, that the accelerated rate of increase of BGT scores cannot be solely or even largely attributed to IQ.

Koppitz (1975) stated that a child of 10 years and 11 months attains a level of developmental maturity at which he or she is able to complete the BGT-*c* without developmentally-based errors. This means that, using the Koppitz Developmental Scoring System, no errors should be made. She adds that an advanced child may reach this ceiling at the age of 8 years. The age of 10 years 11 months corresponds roughly to between 5,5 and 6 years of education and the rate of change of scores in this study accelerates between 6 and 7 years of education. Koppitz (1975) attributed this effect to an age-related developmental maturity in children. However, as already discussed, (see 6.4: p.108) the education effect is implicit in the age of children attending school. Clearly, 6 or more years of education results in the refining of certain aspects of intellectual functioning not evident in less educated adults. It is suggested here that the plateau effect might, in fact, be an educationally-related difference rather than a developmental effect as Koppitz (1975) presumes. However, extensive research is needed for this supposition to be tested.

In the Dyall (1993) study this plateau effect also featured. Dyall (1993) reported findings similar to the present study with the scores of children on the two BGRTs. The rate of change of scores was greatest between children with 3,5 years of education and those with 5,5 years of education, revealing significance on Scheffé's multiple range test at the 1% level for both BGRTs. Scores evened out across the 5,5 and the 8,5 years of education groups, revealing a significance of only 5% on Scheffé's multiple range test for the BGRT-*i* and no significance on the BGRT-*d*. The age means were uniformly stratified and therefore differences in the rate of change in scores between the standards must be attributed to factors other than age differences. Further, the GSAT/NSAGT IQ scores for standard four and seven did not differ significantly while the test results did.

However, the educational level at which the plateau effect in children occurred was somewhat lower than that in the present study. This contrasts with Bender's (1938) theory which maintains that the perceptual ability to organize the gestalt designs increases with maturity and experience. Experience is a time and therefore an age-related factor. Thus low-educated adults could be expected to perform better than children with similar levels of education. However, Hebb (1949) states that effective learning may be determined by early learning. A factor which could account for the scores of children reaching the accelerated rate at a lower educational stage, could be influenced by the quality of schooling, which is discussed in more detail later (see 10.2.5.3.(i): p.186). Adding support to the change of the rate of scores, is Choynowski (1970) who attempts to adjust Koppitz's mean scores using a statistical method of curve fitting. He noted that despite some degree of consistency in scores, "the monotonically decreasing means change in rather irregular ways from one to the next age level" (Choynowski, 1970, p.133). A second possible factor is the grouping of data. The sample in Dyall's (1993) study did not collect data for the educational categories between 3,5 and 5,5 years of education. The present study combines 4-6 years of education in one category. Thus, the exact point of the accelerated rate of change cannot be identified in either study and further research is needed.

10.2.3 Correlations between the three tests of the expanded BGT

The significant correlations between the three BGTs (see Table 8: p.142) have a number of implications which revolve around two important facets, namely the similarities and the differences between the BGT-*c* and the BGRTs.

Firstly, it clearly indicates that there is a degree of overlap in the functions which both types of tests are measuring i.e. the BGT-*c* which is designed to measure visuomotor perceptual organization, and the BGRTs which are a measure of memory. It seems likely that all three tests are, to some extent, measuring a common function, probably that of neurological functioning which relates to visuomotor integration. Clearly this function needs to be intact for adequate performance on the BGT-*c* and the BGRTs.

Secondly, the fact that there is not a higher correlation also suggests that although the tests overlap around 65% in their ability to measure facets of neurological functioning, they are not testing entirely the same type of neurological functioning (see Table 8: p.142). Approximately 35% of the functions measured by each test are different. This is in keeping with the literature reviewed in this study. For example Garron and Cheifetz (1965) stated that the BGRT is more closely related to memory for designs than to the BGT-*c*. It is suggested that the difference between the correlations reported and the highest possible correlation of 1 is due to the fact that while both the BGT-*c* and the BGRTs are measuring brain functioning, they are in fact measuring different aspects of this functioning. As discussed (see 2.5: p.53), it is suggested that while the BGT-*c* measures specific functioning i.e. perceptual visuomotor organization, the memory versions are in fact measuring memory, which is a more diffuse type of functioning.

The fact that the BGT-*c* and the BGRTs are measuring different functions to an extent, supports the rationale of introducing a memory test in addition to the BGT-*c*. It was stated that the combination of the BGRTs and the BGT-*c* render a more sensitive test than either the BGT-*c* or the BGRT on their own (see 2.5: p.53). As earlier stated (see 2.1.5: p.12), Lezak maintains that impairment can be detected by screening tests because they either identify one specific dysfunction or group of dysfunctions, overall dysfunction, or signs that are pathognomic to brain damage (Lezak, 1983). Ensuring that the expanded BGT has all three of these characteristics could increase its sensitivity to brain damage. Also, the hypothesis that the scores from one version may be used to confirm or question the scores of the other version (see 2.5: p.53), is supported, since it appears the BGT-*c* and BGRTs are measuring different functions to some extent.

Little published literature could be found which compares the correlations of the BGT-*c* and the BGRT-*i*. However, Sabatino and Becker (1969) found a correlation of between r=0,61 and r=0,78, significant at the 1% level, between the BGT-*c* and a tachistoscopic administration of the BGT. As already stated (see 2.4.4.1: p.51), research has not finally established that the tachistoscopic version is a measure of memory. Since the BGRT-*d* version administered in this study is new, no literature on this exists. However, the correlations here can be cross-validated by research done on children with similar educational levels (Dyall, 1993). For the correlation between the BGT-*c* and the BGRT-*i* for the educational range of 2,5-9,5 years, a correlation of r=0,68 (p<0,0001) was reported. The correlation between the BGT-*c* and BGRT-*d* was r=0,63 (p<0,0001). The correlation between the BGT-*c* and the sum of the scores of both the BGRTs was r=0,69 (p<0,0001). The results in this study (Dyall, 1993) are very similar to those obtained in the present study (see Table 8: p.142).

The findings of significant differences between the two BGRTs (see 9.3: p.142) support the inclusion of a BGRT-*d*. Literature reviewed (see 2.3.6.2: p.43) revealed that a second, delayed memory test is particularly sensitive to neurological impairment and as discussed earlier (see 10.2.1.1: p.165) the BGRT-*d* appears to be less sensitive to confounding variables such as performance time, extreme care and motivation.

The significant correlations between the BGT-*c* and the BGRTs are consistent with the prediction that when BGT-*c* and BGRT scoring systems are used which are compatible and designed to measure the same criteria, namely brain damage, a correlation between the tests would be evident (see 4.2.7: p.82).

Further, since these correlations are consistent and significant for different groups (see 9.3: p.142), it appears that the relationship between the BGT-*c* and the BRGTs is relatively stable. This means that it is possible that the BGRT scoring system developed in this study has achieved the aim of being compatible with Lacks's (1984) scoring system for the BGT-*c*. However, more research is needed to establish this. This possible compatibility has a number of implications. Firstly, as already stated (see 3.5: p.70) Lacks's (1984) scoring system has a wealth of research testifying to its high diagnostic ability in the identification of brain damage and in the differentiation between functional and organic impairment. The high correlation between the two systems suggests that the rationale used in the development of the BGRT scoring system (see Chapter 7: p.115) is in fact sound and that this method of evaluation is in fact implemental in detecting brain damage.

It also means that a quick, effective method of scoring the BGRT may have been developed. The broad age range to which this system is applicable would obviate the need for the clinician to learn a separate scoring system to test the same function in different age groups. It would also make the BGRT economical and highly practical for the busy clinician both to possess and to learn to administer and score. It was hypothesized (see 2.4.2: p.48 and 4.2.7: p.82) that the limited usage made of the BGRT as well as the failure to compare the BGT-*c* with the BGRT-*i*, could be largely attributed to the lack of a scoring system which was both efficient in identifying brain damage and rapid to score. Perhaps, with this BGRT system, the usage of the BGRT will increase, since these objectives may have been met.

10.2.4 Performance time of the expanded BGT

10.2.4.1 The effect of age on time

Age, in this adult population, exerted no influence on the time taken for the completion of the expanded BGT. However, as already presented (see Table 11: p.145), significant time differences do, occur between the groups. Clearly, a factor other than age, is exerting an influence on the performance time of the expanded BGT.

10.2.4.2 The effect of scores on performance time

With one exception, scores on the BGT-*c* bore no relationship to the performance time (see 9.4.3: p.144). Only the performance time of the 0 years of education group revealed a significant, positive, high correlation to the scores obtained. This finding is highly significant since this group took the longest time to complete this test (see Table 10: p.143). This finding also strongly supports the argument that individuals in this group took more effort to work more carefully (see 10.2.2: p.166). Further, it supports the relationship between time and education, which is discussed more fully in the next section (see 10.2.4.3: p.180). These findings conflict with research cited earlier (see 5.1.4: 94) conducted by Andert *et al.* (1978), who found that the longer the time a subject took to copy the designs on the BGT-*c*, the poorer the protocol was. However, no mention of educational levels were made in these studies and it could be an educational effect which was being revealed here.

Performance time on the BGRT scores only revealed a low correlation which is significant at the 5% level (see 9.4.3: p.144) on scores for the group as a whole, and a moderate correlation of significance for the 0 years of education level. Apart from this, no significant effects of performance time on scores were found. Of interest is that the correlation is negative. This indicates that there is a tendency towards lower scores as performance time increases. This supports the findings of Edwards (1980) who reported that on the BGRT-*i*, scores decreased as the time taken to recall the designs increased, as already discussed (see 5.1.4: p.94). Edwards (1980) made the supposition that subjects were more inclined to make guesses on the BGRT-*i* as performance time increased.

Another point is that the overall trend shows that the correlation coefficients decrease as education increases (see 9.4.3: p.144). This indicates that the time/score correlation might only be applicable to very low-educated individuals. Of note is the fact that the correlations between performance time and scores on the BGT-*c* are almost identical, and on the BGRTs are identical (see 9.4.3: p.144) for the 7-<9 and 9-11 years of education groups. This supports the earlier argument (see 10.2.2.5: p.173) that above 6 years of education for this sample, a plateau is reached where education effects tend to even out.

10.2.4.3 The effect of education on performance time

The highly significant difference in performance time between higher and lower educational levels (see Table 11: p.145) clearly shows that time, at least to a significant degree, is a function of education. As a double check, due to the significance of time in assessing neurological impairment (see 5.1.4: p.94), multiple regression analyses were run and revealed significant correlations between education and performance time, especially on the BGT-*c* (see Table 12: p.147). The fact that the correlation coefficients were all negative indicates an inverse relationship between time and education. Performance time decreases as years of education increase.

No literature to support this was found but this is not surprising since no research on the BGT with low-educated adults could be found. This has strong clinical implications as prior research has connected the presence of brain damage to individuals who take an excessive time to complete the BGT-*c*, as already discussed (see 5.1.4: p.94). For example, Lacks (1984, p.25) stated that performance time exceeding 15 minutes was "usually another strong organic sign". Lezak (1983) also states that slow performance time is linked to neurological impairment (see 2.1.7: p.20). If the findings in the present study are supported by further research, theories regarding performance time which have been developed from studies of more educated adults would be inappropriate and disadvantageous if applied to lower-educated adults.

The issue of time highlights the critical importance of the need for a detailed consideration of administration issues when applying a neuropsychological test to a population to which it has not been previously applied. This consideration is only possible after an extensive review of the literature. If Goodstein *et al.*'s (1955) example of setting a limit for performance time had been followed, the lower-educated subjects would have produced much poorer results, and the importance of the education effect on time would not have been evident.

10.2.4.4 A comparison of performance time of the possibly brain damaged group with the non-clinical group

The lack of significant differences in performance time between the non-clinical and possibly brain damaged groups reaffirm the effects of education on performance time for the expanded BGT. Although only statistically significant for the 0 years of education group, it is of note that for all educational levels, apart from the 7-<9 years of education group, those with possible brain damage completed the BGRT-*d* faster than the non-clinical group.

Two possible reasons for this exist. Firstly, the possibly brain damaged group recalled fewer designs than the normal group, although the difference in recall between the two groups was not significant (see 9.4.4: p.144). It can, however, be hypothesized that since fewer designs were recalled, less time was needed. The second possibility considers an alternative hypothesis. Research reviewed reveals that individuals who have sustained brain damage tire easily and have attentional deficits (see 2.1.8.1: p.22). The addition of the BGRT-*d* lengthens the duration of the test and it is possible that by this time, these subjects were tired and made less effort to recall designs. Further research is needed to firmly establish the reasons for the shorter performance time of the possibly brain damaged group.

The only significant differences for the performance time between the non-clinical and possibly brain damaged groups occur for the group without education on the BGRT-*d*. Once again this supports the indication of an education effect for time, since the lowest education group appears the most sensitive to performance time (see 10.2.4.2: p.179).

10.2.4.5 Time considerations for the group administration of the expanded BGT with low-educated adults regarding Lacks's (1984) option of time as a scorable error

In some cases the time taken to complete the BGT has been regarded as indicative of neurological impairment, as already discussed (see 5.1.4: p.94). By contrast, this study shows no such indication. This finding is supported by Rosecrans and Schaffer (1969) who reported that time as an indicator of cerebral damage was inconclusive. Earlier, the level of education was regarded as "essential" in considering the performance time (see 5.1.4: p.94) and it was hypothesized that low-educated adults would take longer to perform the tests than children with similar levels of education or more educated adults. The findings of this study strongly support this hypothesis.

In the sample of low-educated adults (those with less than 9 years of education including the possibly brain damaged group) only 46 of the 103 subjects completed their BGT-*c* protocol in 15 minutes or less. If Lacks's (1984) optional scoring error of time exceeding 15 minutes had been applied, 55,34% of the sample (57 subjects) would have scored an additional error point. In all, 92,98% of the subjects who would have scored time errors had 6 or less years of education. The addition of one error point would place 12 individuals (11,5%) into Lacks's (1984) neurological range of -5. All of these 12 individuals fell into the 0-3 years of education level. No subjects in the 9-11 years of education group took more than 15 minutes. This again clearly implicates the role of education in performance time as already discussed. The decision taken not to score Lacks's (1984) optional error of time exceeding 15 minutes due to "a likely but unknown" educational effect (see 5.1.4: p.94), has been endorsed as correct by these findings.

In addition to the time considerations discussed above, the fact that the expanded BGT was administered to groups must be appraised. Another reason for not applying Lacks's time error was because of the effect of group testing on performance time (see 5.1.4: p.94), which studies concerning the validity of this administration method (see 2.2.1.3: p.30) fail to note. For example, Koppitz (1975) found that children took a mean time of 6,33 minutes to complete the BGT-*c* while she cites research which found that group administration took approximately 15 to 20 minutes. Dyall (1993) also found that children tested in groups took longer than individually tested subjects. Lacks (1984) reports that non-patient adults took a mean time of 6.01 minutes to complete the BGT-*c* (range: 3-20, *SD*: 2,32) while in this study the most educated group took 10,39 minutes (range: 5-13, *SD*: 5-13). Clearly, performance times for individual testing cannot be applied to the group administration of the BGT.

It was further hypothesized (see 5.1.4: p.94) that the adults in this sample would perform more slowly than children with similar educational levels. Dyall (1993) reported a range of 8-25 minutes on the BGT-*c*. Unfortunately, no further data exists from the study by Dyall (1993) regarding performance time. The performance time range on the BGT-*c* for the non-clinical sample in the present study was 5-45 minutes with a mean time of 18,57 minutes. The differences in the ranges of the two samples clearly indicate that performance time for this sample was longer than for children with similar levels of education. These findings support the suggestions made earlier (see 5.1.4: p.94) that subjects with low education are likely to take longer since it is probable that they are unaccustomed to even holding a pencil or using an eraser. In addition, the anxiety factor, due to the unfamiliar testing environment, was considered and literature reviewed showed that anxiety could slow performance speed (see 5.1.4: p.94).

Before time can be used as a scorable error on the expanded BGT, three considerations need to be clarified by further research with larger numbers of subjects. Firstly, the issue of performance time on the BGT and its relationship to brain damage needs to be reopened and clarified. Secondly, since it is clear that performance time for individual testing differs from that of individual testing, normative data concerning performance time for group testing need to be collected. Finally, because of the strong effect of education on performance time, if time is used as a scorable error following clarification of the neurological impairment issue, norms appropriate to education levels must be applied.

10.2.5 Comparisons and cross validation with other normative data

10.2.5.1 Comparisons of BGT-*c* scores between this study and other studies

(i) Comparisons between this study and Lacks's (1984) norms for adults

The significant difference between Lacks's mean on the BGT-*c* and the mean obtained in this study for individuals with less than 9 years of education (see Table 13: p.148) is explicable considering the education effect on the BGT, as already discussed (see 10.2.2: p.166). Lacks's (1984) sample contains no subjects with less than 3 years education, while the present study does and such subjects consist of 36,6% of the sample. Unfortunately, Lacks (1984) does not provide the mean educational level of these subjects nor their mean age, which would facilitate a closer comparison. Since Lacks's (1984) sample does not include individuals with 0-2 years of education, it could be expected that the mean BGT-*c* score in this study would be significantly lower, which is in fact the case (see Table 13: p.148. The standard deviation of scores in this study is 1.89 and higher than that in Lacks's data which is 1,51. This is consistent with the fact that in this study there is a wider range of low educational levels than in Lacks's (1984) study which would explain the greater variance.

However, for the 9-11 years of education group, the educational level of both studies is more or less the same, as the range is identical and the standard deviations are very similar (see Table 13: p.148). In the present study the standard deviation is 1,36 while Lacks (1984) reports a standard deviation of 1,25 for her sample. Again, Lacks (1984) unfortunately provides no mean for age or educational level. The differences between the means of the two studies for the 9-11 year educational level are not significant and only vary by 0,5, with Lacks's (1984) norm being slightly higher.

The fact that the means for this study and Lacks's (1984) study are so close and reveal no significant differences, shows agreement between the two sets of data. This provides substantial cross-validation of the data obtained in this study. Despite the small sample used in this study, it is clearly consistent with Lacks's (1984) data. All subjects received the same administration of the test and all protocols were marked in the same manner. The above results endorse the validity of this researcher's use and interpretation of Lacks's (1984) scoring system, since for similar education levels, results similar to those of Lacks (1984) were obtained.

(ii) Comparisons between this study and Lacks's (1984) norms for children with similar educational levels

In keeping with expectations (see 6.6: 111), the scores of most of the low-educated adults on the BGT-*c* were significantly lower than those of children with similar educational levels (see Table 13: p.148). As already stated (see 6.6: p.111), the lack of familiarity with pen and paper skills, which school children practice daily, could partially account for this. A second possible explanation is the quality of education which is discussed in more detail later (see 10.2.5.3: p.186). By contrast, the scores of individuals with no education fall between the scores of children of 2,5 and 5,5 years of education. No explanation for this can be found apart from the reasons already given for the performance of this group (see 10.2.2.1: p.166). No data using Lacks's (1984) scoring system with children without education could be found to use as a comparison.

10.2.5.2 Comparisons of BGRT(s) scores between this study and other studies

The only normative data that is in any way available for comparison is Weiss's study (1970) and that of Dyall (1993) obtained for children with education levels similar to the adults in this study. However, Weiss's (1970) mean scores are not entirely satisfactory due to deficiencies in the study (see 4.4: p.83). On observation (see Table 14: p.149), since no statistical analyses could be obtained (see 9.5.2: p.149), the scores on the BGRT-*i* obtained for low-educated adults in this study are very similar to those obtained by Weiss (1970) for children with similar educational levels. By contrast, in the Dyall (1993) study, the only cross-validation with the Weiss (1970) study was for children with approximately 8 years of education. For each educational level, the norms of his study were lower. In other words, there is a similarity between the mean scores obtained in this study and those obtained by Weiss (1970), and both these studies are different from the mean scores of children with similar education levels obtained by Dyall (1993).

One fact which could possibly explain this is the control for normality. In the Dyall (1993) study, very rigorous controls for normality were enforced. No child who had repeated a school class was included in the study. Also, for most of the sample, the researcher had access to the IQ scores and for the entire sample she had access to comprehensive medical records which increased the control for normality. One example indicated this. A child was excluded from the study because of a pre-school epilepsy even although this had not recurred during primary school and the scores of this child fell within the neurological range. In contrast, Weiss (1970) makes no mention of any control for normality whatsoever. In the present study, inherent characteristics in the sample being tested made controlling for normality extremely difficult (see 10.1.3.2: p.156). In addition, certain errors were made in implementing factors which were more possible to control. This is fully discussed later (see 10.3.1: p.194).

Weiss's (1970) scores in all instances were slightly higher than those in the present study, but this can possibly also be explained by the recall scoring system used in this study. Although based on Weiss's (1970) criteria, it was decided that Weiss's (1970) criteria were in fact found these criteria insufficiently described or lenient (see 4.4: p.83), and so the BGRT scoring system used in this research was in fact more stringent than Weiss's (1970), and the scores therefore could be expected to be slightly lower.

In contrast to comparisons with Weiss's (1970) mean scores, the scores in this study differ significantly in most cases with those obtained from children with similar educational levels by Dyall (1993) as Table 15 (see p.151) shows. Similar to the BGT-*c*, there are significant differences between adults and children with similar education levels. The lack of significance between the lowest education groups used in the comparison (see Table 15: p.151) can be explained by the fact that with very limited education, the effect of education might not have not gained sufficient impetus to reveal significant differences.

Apart from the lack of pen and paper skills, as discussed in the previous section in relation to comparisons of BGT-*c* scores, (see 10.2.5.1: p.183), factors which could be responsible for these differences and are discussed in the next section in detail.

10.2.5.3 Factors accounting for the differences between the expanded scores of adults and children with similar educational levels

A number of factors could assist in explaining the difference between the scores of children in the Dyall (1993) study and the adults in the present study with similar levels of education. These involve (i) quality of education, (ii) socioeconomic factors, (iii) sociocultural factors, and (iv) pre-school education.

(i) The quality of education

It has already been established that an educational effect for the BGT exists (see 10.2.2: p.166). South Africa has long been criticized for its schooling of 'other-than-white' individuals. The poor quality of education was highlighted by Viljoen *et al.* (1994, p.208) (see 6.2: p.105). Shochet (1994, p.208) stresses that the role which segregation has played in education is hallmarked by the immense discrepancy in the apportionment of assets between black and white schools. Shochet (1994) stated that such a "disadvantaged social and educational" context cannot be regarded as an indication of the individual's ability. He adds that the same criteria cannot be used to evaluate individuals who are educated in this system and those educated in the more privileged 'traditional white' system. The logical deduction is that such education is not equivalent to that in 'traditionally white' schools.

Regarding the importance of education Crawford-Nutt (1977, p.49) concludes from a review of research that "exposure to education results in increasing differentiation of intellectual structures". In considering the quality of education, Hebb (1949) states that early learning is vital and that the learning of new information is influenced by previous learning. Thus learning may be restricted or enhanced by early learning. In support of this, Crawford-Nutt (1977, p.49) maintains that "formal education is laid upon the foundation of early learning". However, if this early learning is deficient, then it appears from the above that individuals with poor quality education are less able to benefit from later learning. In this sense, Bender (1938) refers to the degree of experience of an individual mediating their ability on the BGT-*c*. Thus, in theory, adults with more experience should fare better than children with similar educational levels. However, in light of the literature above, if the quality of these adults' education was poor, later experience and learning is likely to be limited by a poor quality of education. The effect of a poor quality of education on normative data will be discussed later (see 10.2.7: p.190).

(ii) Sociocultural factors

Sociocultural factors include a lack of familiarity with the testing environment and with pen and paper skills and the cultural aspect regarding this lack of familiarity.

As already stated, the subjects in this sample were likely to be unfamiliar with the testing environment, while school children are regularly exposed to this formal condition. Nell (1994) points out that familiarity with a testing environment can influence normative data. School children regularly perform scholastic tests while the adults in this sample do not. Further, school children also execute neuropsychological tests, and are therefore more familiar than low-educated adults with this situation.

Secondly, while unskilled and semi-skilled low-educated adults are not likely to work with pencils and paper regularly, children do this daily and are therefore more familiar with pen and paper skills and the visuomotor experience of drawing and copying. In addition, the lack of familiarity both with the testing situation and with using pencils and papers could increase anxiety, which in turn could affect scores, as already discussed (see 5.1.2: p.93).

The third sociological factor which could influence performance on the expanded BGT is lack of familiarity. As already discussed (see 6.5: p.110) Yousefi *et al.* (1992) attributed the fact that his norms for Iranian children were lower than the Koppitz means for American children, to factors which included the lack of teaching of art and related visuomotor skills in Iranian schools. This led him to call for the need for normative data appropriate to those who lack such teaching. The current move in South Africa to introduce art education into all schools, and no longer only the 'traditionally white' schools testifies to the lack of such education in the past. Art education could also develop and hone perceptual motor skills which are measured by the BGT. The lack of exposure to the experience of drawing of some of the subjects was succinctly illustrated by the words of a 57 year-old man tested, who said with delight after the Draw-A-Man Test; "I have never drawn a man before".

(iii) Socioeconomic environment

Literature suggests that a lower socioeconomic environment is likely to result in (a) lack of stimulation and (b) undetected brain damage.

Satz & Fletcher (1981) note the importance of environmental factors in contributing to neurological functions from a developmental lag. They state that many individuals in underprivileged environments do not have the necessary stimulation and therefore their recovery rates are below that of more privileged individuals. This emphasizes the need for intellectual stimulation in an individual's development. Liddell and Kvalsvig (1990) note that concomitant with poor educational performance are disadvantaged home situations and unsatisfactory day care. Further, Parsons & Prigatano (1978, p.610) in a discussion of socioeconomic levels, state that factors connected with a low socioeconomic level include decreased "environmental stimulation".

Clearly, the lack of stimulation appears to be a product of low socioeconomic environments and can be prevalent both in a child's early years and in an adult's working environment. Further compounding the problem is that later learning can be limited by early learning (see 10.2.5.1: p.183). School children receive consistent intellectual stimulation. By contrast, the work that low-educated adults do is mostly unskilled or semi-skilled. Therefore, most low-educated individuals are unlikely to receive intellectual stimulation from their daily work environment.

Hargrove and Breazeale (1993) discuss the effects of the decline of rural American communities which appear similar to the problems of underdeveloped rural communities in South Africa. They state that "social and economic threats to rural communities affect health and mental health systems, rendering them progressively less able to respond to critical needs" (Hargrove & Breazeale, 1993 p.319). In addition, they add that such social and economic deterioration is also likely to lead to a decline in the quality of schools. Liddell and Kvalsvig (1990) add that malnutrition coincides with poor education performance. Parsons and Prigatano (1978) support this and state that low socioeconomic levels are associated with deficient health care at the time of birth and during childhood. Such poor medical care can lead to undetected neurological damage.

To summarize, a low socioeconomic level is related to poor physical conditions and insufficient stimulation. Poor physical conditions include poor health care and malnutrition, both of which can increase the risk of neurological impairment, albeit undetected. Lack of stimulation and a poor quality education can inhibit the development and differentiation of intellectual functioning. These factors assist in explaining why children in 'traditionally white' schools fared better on the expanded BGT, than adults with similar levels of education. Again, the need for specific normative data is highlighted.

10.2.6 Performance of older adults

The findings in this study that more older people from a normative sample score within the range of neurological impairment than younger subjects is in keeping with literature already reviewed (see 6.4: p.108). It appears that those above 59 years of age were correctly excluded from the non-clinical sample. Although the sample of older people was small, it shows a trend consistent with that of Lacks and Storandt (1982) who found that 22% of individuals aged 60-74 years old scored within the organic range of the BGT. The fact that 80% scored within the organic range creates a new research question since it appears that both age and education effects resulted in these poor scores. However, the sample in this study is far too small to reach any conclusions, but it does indicate an area of research for the future. Contrary to the findings of Youngjohn and Crook III (1993) already cited (see 2.3.6.2: p.43), older adults did not recall more after a delay. However the authors added that their sample had a high educational level and that their findings might not apply to lower-educated individuals.

10.2.7 The need for appropriate normative data

The need for normative data on neuropsychological tests which are current and appropriate to the populations to which it is applied is unequivocally supported by literature (e.g. Adams *et al.*, 1982; Anastasi, 1976; Bornstein, 1986; Brook, 1975; Choynowski, 1970; Crawford-Nutt, 1977; Hart, 1989; Leckliter & Matarazzo, 1989; Loewenstein *et al.*, 1994; Satz & Fletcher, 1981; Shochet, 1994; Yousefi *et al.*, 1992; Zuelzer, 1976 (see 6.5: p.110).

This research set out to investigate the effect of little or no education on BGT scores (see Chapter 1, p.1), since literature was reviewed which showed that educational levels can effect the scores of neurospychological tests in general, scores on the BGT-*c*, and also scores on memory tests (see 2.2.2.2: p.34). However, the effect of the *quality* of education and sociocultural and socioeconomic factors have also emerged in the course of this study (see 10.2.5.3: p.186).

Thus, the results of this research highlight two inter-related population groups for which separate normative data is needed on the BGT. Firstly, normative data for individuals with little or no education is necessary as already discussed (see 10.2.2: p.166), as BGT scores are generally lower for individuals in the sample of adults tested in this research who have less education. Secondly, normative data appropriate to the quality of education and sociocultural and socioeconomic conditions is also needed.

The effect of low educational levels has been comprehensively covered in this research (see 10.2.2: p.166) and no further discussion is needed. However, since the quality of education was not considered from the start, this issue needs to be appraised in more detail, with reference to literature. This concurs with earlier statements that increased knowledge and information leads to an increased differentiation of questions (see Preface: p.xvii).

Loewenstein *et al.*'s (1994, p.624) studies in Latin America highlighted both the paucity of and need for normative data for particular culturally dissimilar populations. Zuelzer *et al.* (1976) and Fisher (1993) support this. Further, Brook (1975) states that normative values alter with time because of (i) improvements in education which in turn results in further development, and (ii) the natural progression of development of the human being across time. Nell (1994) supports this and states that norms which were established 40 years ago are obsolete. Further, Nell (1994, p.103) attributes the higher normative data appropriate to Americans than to South Africans on the same neuropsychological test, to the fact that Americans have become "both better educated and 'test-wise'". He adds that the more outdated the norms, the more individuals with deficits will score within the non-organic range while those without impairment will score increasingly higher. Since Lack's (1984) norms are relatively current, this aging of normative data is not relevant to this study. However, the *reason* for the dating of normative data is highly pertinent. Nell (1994, p.103) sums up the argument by stating "old norms do not reflect the ongoing educational and experiential development" of individuals.

The importance of the quality of schooling has already been discussed (see 10.2.5.3(i): p.186). If improvements in education lead to increased development and higher test scores, it can be expected that a disparity in development exists between those who have been subjected to inferior eduction and those who have benefitted from a more advanced educational system. Thus, scores on neuropsychological tests for educationally-disadvantaged individuals are likely to be lower than for more privileged individuals. The results of this study endorse this supposition (see 10.2.5: p.183).

Nell (1994) provides a comprehensive discussion of conflicting tensions in South Africa regarding separate normative data for separate racial and cultural groups. On the one hand, psychologists do not want to be accused of racism, yet on the other hand, there is a need for normative data for separate racial and cultural groups in the country. Regarding assessment Nell (1994, p.107)) states:

> The idea of separate norms for separate groups is deeply repugnant to South Africans; claims for human universals ... are more comfortable, and

> politically more correct ... any recognition of race and cultures as modifiers of human performance thus smacks of racism. Cultural universalism was uncritically accepted as an alternative view. It became very difficult for 'progressive' South African psychologists to recognize that an acceptance of cultural variation as a moderator of test scores, and the consideration of such scores in the context of cultural difference, as advocated by crosscultural psychology, is an inherently kinder view...
>
> In order to deal vigorously with the legacy of apartheid, South African psychologists will have to face the reality of race as a mediator of the quality of education, economic opportunity ... and the advantagement-disadvantagement. This complex of issues cannot be addressed by hiding behind 'inoffensive' terms such as home language or culture.

Nell (1994, p.107) concludes that norms which are developed without prior construct validation are not norms, but a "numeric fiction".

The dilemma brought about by the need for separate normative data is not a specifically South African problem. Loewenstein *et al.* (1984, pp.626, 629) precisely defines the choice of universalism or separatism which the neuropsychology discipline is facing:

> We are presently at a crossroads ... where we can decide to increase the precision of our science to better incorporate the diversity of human beings in our population, or we can maintain the *status quo.*
>
> Psychology has traditionally prided itself on the respect for individual differences and identification of those factors that influence behaviour ... we owe it to our patients and to our profession to offer the very best technology of which we are capable.

It is clear that due to the poor schooling quality and other sociocultural and socioeconomic factors, individuals in these groups score lower on neuropsychological tests and therefore separate normative data needs to be developed. The fact that many of the individuals in low educational and socioeconomic groups are 'other-than-white' is a product of the past era of apartheid, and reflects neither a racial bias in the BGT nor the attitude of this author. On the contrary, failure to acknowledge the existence of lower scores for these groups will perpetuate the disadvantages experienced by these groups in the past.

In addition, it is also clear that education plays a vital role in the performance of neuropsychological tests in general and on the BGT in specific. Due to the large numbers in this population, normative data applicable to low educational levels and those who have received a poor quality education are urgently needed.

10.2.8 Preliminary normative data

This study presents preliminary normative data for low-educated adults. It is stressed that small sample numbers were used and more research with larger numbers is needed to establish more representative normative data. However, the close correlations with Lacks's (1984) scores for a similar educational group, as already discussed, do support the accuracy of the scores in this study. In addition the BGT-*c* scores for the 0 years of education group are unlikely to be indicative of those in a normal population distribution, for reasons already discussed (see 10.2.2: p.166)

Regarding the applicability of these data to other racial groups, Dyall (1993) found a parity of expanded BGT scores across races, among children attending 'traditionally-white' schools. This lack of racial differences remained constant even although significant IQ differences existed which were negatively biased for 'other-than-white' subjects. Thus it appears that no racial bias exists on the expanded BGT when it is evaluated according to the scoring systems used in this study. While it is realised that the data presented here is only a small sample, it can provide a diagnostic guideline, which is useful as no such other data exists in any published studies. Considering the lack of race effects, the normative data collected in this study, could, at the very least, be used confidently as a rough guideline amongst other racial groups with low education. No other norms specifically stratified for low education levels exist in published research to the author's knowledge. BGT scores, when using the scoring systems applied to this research, appear not to have a race effect (Dyall, 1993) and thus provide a relatively accurate, racially non-biased assessment. In contrast, significant educational effects have been found on this test. For this reason and as discussed (see 6.5: p.110) it would seem more appropriate to apply the norms in this study to low-educated individuals of other races than to apply culturally correct norms which do not account for a low educational level.

It is also important to note, that for both clinical or research use, the administration procedures used in this study should be strictly adhered to for application of or comparison with the data presented here. As stressed throughout this thesis, much research on the BGT is contradictory as different procedures have been followed. Changes in procedures, especially factors such as observation and scoring a time error, are likely to influence results.

10.3 Weaknesses of the study

The selection of subjects in this study had flaws. Some of these would have been possible to control and constitute weaknesses in this study. Others were inherent in the sample group tested and reflect the problems with testing low-educated individuals.

10.3.1 Controllable weaknesses of this study

Two possible confounding variables were not taken into account in this study. Firstly, subject selection did not control for alcohol abuse. However, it would have been difficult to determine the level of alcohol consumption at which subjects were excluded. Secondly, a weakness existed in the establishment of subjects' educational levels. Participants were not asked if they were attending or had attended a literacy course, which could make their level of education higher than the last standard which they passed. Also, no consideration was given to the fact that an applicant could have spent more than one year in the same standard. Although a subject repeating a standard would still have the same educational level as a subject who passed the first time, a subject repeating a standard would gain an additional year's experience of perceptual visuomotor, pen and paper and other skills.

While the above weaknesses are noted, it must be pointed out that many researchers have failed to implement even an attempt to control for normality both internationally (e.g. Weiss, 1970) and in South Africa (Viljoen *et al.*, 1994). In these studies all available subjects were tested without any effort by the researchers to exclude any individuals with possible neurological impairment.

10.3.2 Inherent weaknesses of this study

Certain problems in controlling for normality were inherent features of either the method of recruiting subjects or of the subjects themselves and included the honesty of subjects, the quality of their education and their knowledge of their own medical history.

Since the tester was known to have contact with the employer, there may have been individuals who did in fact smoke more than three reefers of cannabis a day or who had sustained neurological damage, but who did not report it for fear of vocational repercussions. The fact that some subjects did acknowledge the above conditions means that in some cases at least, the tester obtained the trust of the subjects. Secondly, a more difficult problem concerns the *quality* of education, which has already been fully discussed (see 10.2.5.3(i): p.186).

The third, and most serious threat to the validity of this study is that it is unlikely that low-educated adults have a comprehensive knowledge of their own medical history and therefore subjects who had suffered neurologically-implicated diseases might have been included. Two circumstances which arose during testing illustrated this. A number of subjects did not know their age, but some knew their date of birth or produced an identity document. Also, as already mentioned (see 10.1.2.2: 154) three subjects who had epilepsy were not aware of its medical name.

Educated adults generally have a good idea of their own medical histories, since this information is accessible to recording in the written form. For example, 'traditionally white' schools keep comprehensive medical records of each child which are updated constantly. In a sample such as the one above, a number of possibilities could prevent subjects from having a knowledge of their medical history. The parents of those with illness during childhood might not have been told the name of the illness or might not have been able to recall it. Even if they knew and remembered the name of the illness they might not have realised the importance of passing this on to their child. Alternatively, subjects hospitalized as adults might not have been told the name of their illness or might not have remembered it. Further, considering the language differences and earlier poor quality of medical care for the 'other-than-white' groups in this country, there is the possibility that meningitis and similar illness were not diagnosed. Additionally, considering the inaccessibility of health care in the Eastern Cape, a subject might well have had meningitis and not even visited a hospital, due to ignorance, lack of funds or distance. Another possibility is that an ill subject could quite possibly have consulted a traditional healer rather than a western medical doctor. Further, Bornstein (1986) states that it is possible that in a sample of 'normal' individuals, some subjects might have neurological impairment which might be unknown to the individual. Cases which would not be identified by screening questions include pre-birth dysfunctions. Thus subjects with neurological damage which was undetected or unknown to them could well have been included in the non-clinical sample.

10.4 The processing of scores on the expanded BGTs

10.4.1 Conversion of data to a standard score

Scoring the BGRTs according to the recall scoring system presented in this research (see Chapter 7: p.115) creates a possible compatibility between the BGT-*c* and BGRTs (see 10.2.3: p.176). This is because the criteria used to evaluate errors on the BGT-*c* are similar to and based on the same principles, as those used to determine the correctness of recalled designs. Thus the scores of the BGT-*c* and the BGRTs may now, in theory, be meaningfully compared. However, a mathematically correct manner of comparing the scores is necessary to maximize the full potential of the data obtained from the expanded BGT.

In this study statistical correlations were used to establish the relationship between the tests. While this is appropriate to research, it is not practical in everyday assessment situations. Firstly, it is often desirable to make an immediate evaluation, yet this would depend on the availability of a computer with which to evaluate the data. Secondly, psychologists in training must acquire a vast expanse of knowledge in many areas and statistics are seldom given a high priority in clinical Masters' training programmes. Even where such training is received, the nature of statistics is such that when not frequently used, individuals are inclined to forget the application thereof. Thus, for optimal use of the expanded BGT, a quick and simple method of meaningfully comparing the scores on the BGT-*c* and the BGRTs must be found. For the reasons stated, this method should not require the use of a computer or complicated statistical procedures.

Lacks's (1984) scores are error scores and therefore negative, while the BGRT scores are positive. A direct comparison between the BGT-*c* and the BGRTs would therefore involve an inverse relationship. To enable a direct comparison between the two sets of scores, Lacks's (1984) scores can be converted to positive correlates (standard scores) equivalent to Lacks's (1984) raw scores. Table 16 below provides each Lacks's (1984) raw score with its equivalent standard score.

Using this conversion, the standard score equivalent to a perfect Lacks's (1984) raw score of 0 (for a protocol without errors) is 12 - i.e., the highest score on the standard conversion table. Similarly, the poorest Lacks (1984) score of -12, has a standard score of zero i.e., the lowest possible score on the standard scoring scale (see Table 16: p.198).

10.4.2 Equalizing the weights on the BGT-*c* and BGRT scoring systems by converting to percentages

A further hindrance to a direct comparison between the standard BGT-*c* scores and those of the BGRTs is that the number of criteria in each system differ. Lacks's (1984) BGT-*c* scoring system has 12 items while the BGRT scoring system has 8 items. Thus, a direct comparison would be unbalanced as each system has a different weight. This can be overcome by equalizing the weights of both systems, which is effected by converting all scores into percentages. A percentage for each possible score is obtained by dividing the score by the number of items in the scoring system and then multiplying by 100. This procedure in indicated in Table 17 below.

These scores may now be directly compared. For example, a raw score of -3 on the BGT-*c* of the test is equivalent to a standard score of 9 which results in a percentage of 75%. A score of 5 on the BGRTs has a percentage of 62,5%.

10.4.3 Equalizing the memory component of the expanded BGT

The last problem to consider is that one test measures perceptual visuomotor perception and two tests reflect memory. For the copying and the memory components to carry equal weight, the BGT-*c* should be regarded as one unit and the two memory components as a second, single unit. By dividing the scores of the two BGRTs by half and summing them up, the performance of visual perceptual motor skills (BGT-*c*) may be directly and equally compared with the measurement of memory (BGRT-*i* and BGRT-*d*). This is shown in Table 19 below.

Table 16: The conversion of Lacks's (1984) raw scores to standard scores for the BGT-*c*

Raw Score	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
Standard Score	0	1	2	3	4	5	6	7	8	9	10	11	12

Table 17: The equalization of the BGT-*c* and BGRTs scores to percentages

Standard score divided by number of items	$\frac{0}{12}$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{4}{12}$	$\frac{5}{12}$	$\frac{6}{12}$	$\frac{7}{12}$	$\frac{8}{12}$	$\frac{9}{12}$	$\frac{10}{12}$	$\frac{11}{12}$	$\frac{12}{12}$
Decimal	0	0,11	0,16	0,25	0,33	0,41	0,50	0,58	0,66	0,75	0,83	0,91	1
Percentage	0	11,11	16,66	25	33.38	41,66	50	58,33	66,66	75	83,33	91,66	100

Table 18: The equalization of the BGRTs scores to percentages

Raw score divided by number of items	$\frac{0}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{5}{8}$	$\frac{6}{8}$	$\frac{7}{8}$	$\frac{8}{8}$
Decimal	0	0,125	0,25	0,375	0,05	0,625	0,75	0,875	1,00
Percentage	0	12,5	25,00	37,50	50,00	62,50	75,00	87,50	100,00

10.4.4 Combining the conversion processes

The information from the tables already presented can be combined into a single set of tables to facilitate conversion. Table 20 and Table 21 below provide the combination of the tables presented in the rationale of this mathematical model and facilitate quick conversions. While the procedure is fairly lengthy to explicate, the actual operations involved will take no more than a minute to perform, with or without a calculator, using the tables provided. A further advantage of this system is that in instances in which large samples of data need to be analyzed, as in the case of research samples, a very simple computer programme can be designed to execute this rapidly.

Now the BGT-*c* and BGRT scores can act as supplementary information for each other, and act as a guard against misdiagnosis. For example, Lacks (1984) regards a raw score of -5 for the BGT-*c* as indicative of brain damage. The conversion of this raw score, according to the process described, would result in a standard score of 7 and a percentage of 58,33%. This may be directly compared with the BGRT percentage. A raw recall score of 7 on the BGRT-*i* Test and 6 on the BGRT-*d* would lead to a memory percentage of 87,5% and 75% respectively and an overall memory percentage of 81,25%.

When this BGRT percentage is compared with the BGT-*c* percentage, the clinician is alerted to the possibility of a false positive diagnosis and can conduct further tests to clarify the diagnosis. Similarly, a Lacks (1984) raw score of -4 (66,66%) can be a borderline case, particularly if the execution of the protocol took a long time. As already discussed (see 5.1.4: p.94), Lacks (1984) states that with enough time, organically impaired individuals may manage to execute a protocol that is borderline rather than indicative of brain damage. If the memory component was, for example, a raw score of 1 for the BGRT-*i* version and 0 for the BGRT-*d* version, producing a final memory percentage of 11,11%, the clinician is cautioned as to the possibility of a false negative. Again, further testing would be indicated in such a situation.

Table 19: The equalizaton of scores from raw scores to percentages divided by two, to equalize the weight of the two BGRTs

Standard score divided by number of items	0/8	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
Decimal	0	0,125	0,25	0,375	0,05	0,625	0,75	0,875	1,00
Percentage	0	12,5	25,00	37,50	50,00	62,50	75,00	87,50	100,00
Percentage/2	0	6,25	12,5	18,75	25	31,25	37,5	43,75	50

Table 20: The conversion of Lacks's (1984) raw scores for the BGT-*c* to equally weighted BGT quotient scores

Raw Score	-12	-11	-10	--9	-8	-7	-6	-5	-4	-3	-2	-1	0
Standard Score	0	1	2	3	4	5	6	7	8	9	10	11	12
Standard score divided by number of items	0/12	1/12	2/12	3/12	4/12	5/12	6/12	7/12	8/12	9/12	10/12	11/12	12/12
Decimal	0	0,11	0,16	0,25	0,33	0,41	0,50	0,58	0,66	0,75	0,83	0,91	1
Percentage	0	11,11	16,66	25	33,33	41,66	50	58,33	66,66	75	83,33	91,66	100

Table 21: The conversion of Lacks's (1984) raw scores for the BRGTs to equally weighted BGT quotient scores

Raw score divided by number of items	0/8	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
Decimal	0	0,125	0,25	0,375	0,05	0,625	0,75	0,875	1,00
Percentage	0	12,5	25,00	37,50	50,00	62,50	75,00	87,50	100,00

10.4.5 Creating a single quotient for the expanded BGTs

Finally, performance on the BGTs can be reduced to a single BGT quotient. As the two percentages now carry equal weight they can be combined in a meaningful way. Adding the two percentages scores together and dividing by two, to return to a percentage, will produce a final BGT quotient. An equal weight of 50% significance is now distributed to the two functions being measured, namely visual motor perception and memory. For maximum information, both the individual percentages and the BG quotient should be known. It is therefore suggested that the final scores be displayed as follows:

66,66% + 11,11% (11,11% + 0%) = 77,77/2 = 38,86%

BGT-*c* + BGRTs (BGRT-*i* + BGRT-*d*) = BGT quotient/2 = BGT quotient

The first percentage indicates the BGT-*c* percentage, the second the memory percentage with the individual percentages of the memory tests indicated in parenthesis. The sums of both percentages are divided by two to reduce the score back to a percentage again. A cautionary note is necessary here. As already discussed (see 10.2.2.2: p.168), it appears that the BGT-*d* is, at least partly, detecting a different type of memory functioning to that identified by the BGRT-*i*. Thus, it is important for accuracy that the individual scores on both memory components, the BGRT-*i* and the BGRT-*d*, are retained as indicated above.

Using this method, a very low percentage quotient, such as 20%, would clearly be indicative of brain damage and a very high percentage quotient, such as 90%, would mean the individual being testing is highly unlikely to have neurological impairment. It is already known that a Lacks (1984) raw score of -5 is indicative of neurological impairment. This is equivalent to 58,33%. However, further research is needed to establish at what precise point the memory component percentage could predict brain damage with an acceptably high degree of accuracy. Once this is known, then a cut-off point for the BGT quotient can also be established. Further, in order to be diagnostically useful, normative data which are stratified according to age and education levels are necessary for the BGT-*c*, BGRT-*i* and BGRT-*d* versions of the expanded BGT. The conversion of the preliminary normative data for adults with less than 9 years of education, already presented in this research (see Table 4: 138), is provided in Table 22. As can be seen, these percentage scores are easier to work with than the equivalent data presented earlier in Table 4 (see p.138).

It was stated earlier that this thesis would argue that the expanded BGT would meet all Lezak's (1983) criteria for screening tests (see 2.1.5: p.12). In this section, Lezak (1983) was cited as stating that impairment can be detected by screening tests because they either identify one specific dysfunction or group of dysfunctions, overall dysfunction, or signs that are pathognomic to brain damage. Specific functions pertaining to visuomotor perceptual functioning are measured by the BGT-*c* (see 2.1.7: p.20). The BGRTs measure memory, which as already stated (see 2.4: p.46) is indicative of diffuse impairment. Finally, both the scoring systems penalize features which Lezak (1983) regards as 'diagnostic signs' such as perseveration and rotation. This contributes another, inherent indicator of brain damage in the expanded BGT. Thus all three of Lezak's (1983) criteria for screening tests are measured by the expanded BGT. The conversion discussed above allows these functions to be easily and rapidly interpreted. While the individual components in the final BGT quotient indicated the different aspects of neurological functioning, the BGT quotient itself indicates the combination of these functions.

Table 22: The processing of scores on the expanded BGT to accord equal weight to the BGT-*c* and BGRTs and to produce an expanded BGT quotient (all figures represent percentages)

Educational level (years)	BGT	$\frac{\text{BGRT-}i + \text{-}d}{2}$	BGRT-*i*	BGRT-*d*	Expanded BGT quotient
0	66,89	19,27	25,19	13,34	43,08
1 - 3	53,89	18,52	20,74	16,30	36,21
4 - 6	60,09	24,06	28,81	19,30	42,08
7 - <9	73,99	39,73	42,09	37,37	56,86
Sub-total	**66,16**	**27,78**	**30,62**	**24,93**	**46,97**
9 - 11	76,92	40,17	41,88	38,46	58,55
TOTAL	**67,63**	**29,47**	**32,16**	**26,78**	**48,55**

CHAPTER 11: CONCLUSION AND RECOMMENDATIONS

The aims of this study, as presented in Chapter 1, have been met. Firstly, the expanded BGT appears to be a more sensitive measure than the BGT-*c* test alone. In addition, the significant difference between the BGRT-*i* and the BGRT-*d* suggests that the addition of a delayed recall enhances the visual memory functions measured by the expanded BGT. This fulfils the first aim of this study.

Secondly, administration procedures for the group testing of low-educated adults were developed and described in this research. Other administration-related decisions were also addressed. This accomplishes the second aim of this study as the group testing method renders the expanded BGT economical and rapid to apply and the description of procedures allows for replication of the method used.

Further, a rapid scoring system for the BGRTs has been clearly operationalized in order to make it accessible to clinicians and researchers. The high correlation between the BGT-*c* scores and the BGRT's using the recall scoring system developed, indicates that these two systems are compatible and a direct comparison of results is possible between the BGT-*c* and BGRTs. The conversion of data, using the method suggested in 10.4 (p.196), as suggested in this thesis, enables equal weight to be attached to both the versions in a manner which is rapid and simple. Thus, aims three and four have been accomplished.

The fifth aim was achieved with the investigation of educational effects with regard to scores and performance time of low-educated adults on the expanded BGT. As predicted, these appear substantial. Finally, this research has presented preliminary normative data of the scores and performance time on the expanded BGT for low-educated adults using Lacks's (1984) scoring system on the BGT-*c*, and the system developed in this research to score the BGRTs. This addressed the sixth and final aim of this study as presented in Chapter 1 (p.1).

The procedures and scoring system developed for the BGRT means that this test can now be used with ease and confidence. Hopefully, this will once again stimulate the use of the test and promote further research into its abilities. The compatibility of the two scoring systems will facilitate additional data, clinicians as opposed to researchers applying the BGRT instead of the BGT-*c*.

In addition, secondary aims were met with the development of a group administration method suitable for low-educated and uneducated adults. Other administration-related problems were also addressed. Hopefully, by drawing attention to research inaccuracies, lack of standardization and errors, future researchers will be more aware of the variety of issues on this 'deceptively' simple test.

The value of any research would be decreased if it revealed no problems which could not be solved within the scope of the study. In this study, a number of new research questions have emerged. The issue of performance time is probably the most important factor which needs to be addressed by future research. The findings of this study show that education is inversely related to performance time, with the least educated individuals taking the longest time to complete the expanded BGT. The time issue is particularly pertinent as some researchers in the past have regarded a long performance time as an indication of neurological impairment. Further research needs to establish whether performance time is related to education alone, or to both education and neurological impairment. It may also be that the education effect on performance time is only applicable to those with low levels of education or no education. In addition, the performance times for non-clinical adults on group testing need to be collected. Until such research is available, it seems inappropriate to apply an error score for performance time when administering the BGT in any form, especially with low-educated adults.

Also of importance is the need for research which establishes the neurological cut-off score for low-educated adults on the BGT-*c*. Along with this, cut-off scores for the two BGRT's and for the test as a whole need to be established. Also, scores obtained by low-educated adults who are neurologically or functionally impaired need to be collected. In addition, comprehensive normative data for this population group using larger samples needs to be established.

Another applicable research question surrounds the performance of the 0 years of education group. While possible hypotheses were raised in the discussion above, more formal research is necessary to elucidate the scores obtained by this group. The pertinent question is will low-educated adults still score less than children with similar educational levels, if the quality of education is similar?

In addition, the scoring of rotations, as already discussed needs to be thoroughly investigated. The plateau effect seen in this study is also of interest. Research is needed to establish whether this is peculiar to this research, specific to the BGT or whether it a trend with other neuropsychological tests. A secondary question which is raised is the incidence of neurological impairment among the Eastern Cape 'other-than-white' population, the reasons for this and implications thereof.

In conclusion, researchers and clinicians should become more aware of the influence of education on the BGT and possibly on other neuropsychological tests. The continued use of norms for well-educated adults inappropriately applied to low-educated individuals will perpetuate the disadvantages experienced by this population and lead to diagnostic errors.

REFERENCES

Aaronson, B.S. (1957). The Porteus Mazes and Bender Gestalt Recall. *Journal of Clinical Psychology, 13*, 186-187.

Adams, R.L., Boake, C. & Crain, C. (1982). Bias in a neuropsychological test classification related to education, age, and ethnicity. *Journal of Consulting and Clinical Psychology, 50*(1), 143-145.

Aikman, K.G., Belter, R.W. & Finch Jnr, A.J. (1992). Human figure drawings: A validity in assessing intellectual level and academic achievement. *Journal of Clinical Psychology, 48*(1), 114-120.

Allen, R.M. & Frank, G.H. (1963). Experimental variations of the mode of reproduction of the Bender Gestalt stimuli. *Journal of Clinical Psychology, 19*, 212-214.

Allen, R.M., Adamo, C., Alker, L.N. & Levine, M.N. (1971). A study of recognition and reproduction of Bender Gestalt figures by children of average and below intelligence. *The Journal of Genetic Psychology, 119*, 75-78.

American Psychological Association (APA) (1994). *Diagnostic and statistical menual of mental disorders* (4th ed.). Washington D.C.: American Psychiatric Press.

Anastasi, A. (1976). *Psychological testing* (4th ed.). USA: Macmillan.

Anderson, B. & Rallis, K. (1981). Relationship between Bender errors, emotional indicators and performance on Bender Recall. *Perceptual and Motor Skills, 53*, 497-498.

Andert, J.N., Hustak, T.L. & Dinning, W.D. (1978). Bender-Gestalt reproduction times for retarded adults. *Journal of Clinical Psychology, 34*(4), 927-929.

Anthony, W.Z., Heaton, R.K. & Lehman, R.A.W. (1988). An attempt to cross-validate two actuarial systems for neuropsychological test interpretation. *Journal of Consulting and Clinical Psychology*, 48, 317-326.

Arbit, J. & Zager, R. (1978). Psychometrics of a neuropsychological test battery. *Journal of Clinical Psychology, 34*(2), 460-465.

Armentrout, J.A. (1976). Bender Gestalt recall: Memory measure or intelligence estimate? *Journal of Clinical Psychology, 32*(4), 832-834.

Armstrong, R.G. (1965). A Re-evaluation of copied and recalled Bender-Gestalt reproductions. *Journal of Projective Techniques, 29*, 134-139.

Arnold, L.E., Huestis, R.D., Wemmer, D. & Smeltzer, D.J. (1978). Differential effect of amphetamine optical isomers on Bender Gestalt Performance of the minimally brain dysfunctioned. *Journal of Learning Disabilities, 11*(3), 127-132.

Bach, Z. (1987). Keep "Gestalt" in the name of the test: The curious name permutations of the Bender-Gestalt Test. *Journal of Personality Assessment, 51*(1), 109-111.

Baddeley, A. (1987). *Working memory*. New York: Oxford University Press.

Barrett Jnr, E.T., Wheatley, R.D. & Laplant, R.J. (1982). A brief clinical neuropsychologic screening battery: Statistical classification trials. *Journal of Clinical Psychology, 38*(2), 375-377.

Becker, J.T. & Sabatino, D.A. (1971). Reliability of individual tests of perception administered utilizing group techniques. *Journal of Clinical Psychology, 27*, 86-88.

Bender, L. (1938). *A Visual Motor Gestalt Test and its clinical use*. New York: American Orthopsychiatric Association.

Bender, L. (1946). *Instructions for the use of Visual Motor Gestalt Test*. New York: The American Orthopsychiatric Association.

Bender, L. (1965). On the proper use of Bender Gestalt test. *Perceptual and Motor Skill s, 20*, 189-190.

Bigler, E.D. & Ehrfurth, J.W. (1981). The continued inappropriate singular use of the Bender Visual Motor Gestalt Test. *Professional Psychology, 12*(5), 562-569.

Billingslea, F.Y. (1963). The Bender Gestalt: A review and perspective. *Psychological Bulletin, 60*(3), 233-251.

Binder, L.M. & Rattok, J. (1989). Assessment of the postconcussive syndrome after mild head trauma. In M.D. Lezak (Ed.), *Assessment of the behavioural consequences of head trauma* (pp.37-48). New York: Liss.

Black, F.W. (1973). Reversal and rotation errors by normal and retarded readers. *Perceptual and Motor Skills, 36*, 895-898.

Boake, C.T. & Adams, R.L. (1982). Clinical utility of the background interference procedure for the Bender-Gestalt test. *Journal of Clinical Psychology, 38*(3), 627-631.

Boll, T.J. (1974). Psychological differentiation of patients with schizophrenia versus lateralized cerebrovascular, neoplastic, or traumatic brain damage. *Journal of Abnormal Psychology, 83*(4), 458, 83(4), 456-458.

Bornstein, R.A. (1986). Classification rates obtained with "standard" cut-off scores on selected neuropsychological measures. *Journal of Clinical and Experimental Neuropsychology, 8*(4), 413-420.

Bowland, J.A. & Deabler, H.L. (1956). A Bender-Gestalt diagnostic validity study. *Journal of Clinical Psychology, 12*, 82-84.

Brannigan, G.G. & Brunner, N.A. (1991). Relationship between two scoring systems for the modified version of the Bender-Gestalt Test. *Perceptual and Motor Skills, 72*, 286.

Brilliant, P.J. & Gynther, M.D. (1963). Relationships between performance on three tests for organicity and selected patient variables. *Journal of Consulting Psychology, 27*(6), 474-479.

Brook, R.M. (1975). Visual Retention Test: Local norms and impact of short-term memory. *Perceptual and Motor Skills, 40*, 967-970.

Brooks, N. (1989). Close head trauma: Assessing the common cognitive problems. In M.D. Lezak (Ed.), *Assessment of the behavioural consequences of head trauma* (pp.61-85). New York: Liss.

Bruhn, A.R. & Reed, M.R. (1975). Simulation of brain damage on the Bender-Gestalt Test by college subjects. *Journal of Personality Assessment, 39*, 244-255.

Burgess, M.M., Kodanaz, A., Ziegler, D. & Greenburg, H. (1970). Prediction of brain damage in two clinical populations. *Perceptual and Motor Skills, 30*, 523-532.

Butler, O.T., Coursey, R.D. & Gatz, M. (1976). Comparison of the Bender Gestalt Test for both black and white brain-damaged patients using two scoring systems. *Journal of Consulting and Clinical Psychology, 44*(2), 289-285.

Canter, A. (1966). A background interference procedure to increase sensitivity of the Bender-Gestalt Test to organic brain disorder. *Journal of Consulting Psychology, 30*(2), 91-97.

Carlson, L.D. (1966). A comparison of Negro and Caucasian performances on the Bender-Gestalt Test. *Journal of Clinical Psychology, 22*, 96-98.

Cerbus, G. & Oziel, L.J. (1971). Correlation of the Bender-Gestalt and WISC for Negro children. *Perceptual and Motor Skills, 32*, 276.

Chavez, E.L., Trautt, G.M., Brandon, A. & Steyaert, J. (1983). Effects of test anxiety and sex of subject on neuropsychological test performance: Finger Tapping, Trail Making, Digit Span and Digit Symbol Tests. *Perceptual and Motor Skills, 36*, 923-929.

Chouinard, M-J. & Braun, C.M. (1993). A meta-analysis of the relative sensitivity of neuropsychological screening tests. *Journal of Clinical and Experimental Neuropsychology, 15*(4), 591-607.

Choynowski, M. (1970). Curve-fitting as a method of statistical correction of developmental norms, shown on the example of the Bender-Koppitz test. *Journal of Clinical Psychology, 26*, 133-141.

Clark, R.E.H. (1982). The identification of the learning disabled child through the use of the Bender Visual Motor Gestalt Test as a measure of visual memory (Doctoral dissertation, University of Oklahoma, 1981). *Dissertation Abstracts International, 42A*(7), 3103.

Clarke, B.R. & Leslie, P.T. (1971). Visual-motor skills and reading ability of deaf children. *Perceptual and Motor Skills, 33*, 263-268.

Coleman, J.C., Butcher, J.N. & Carson, R.C. (1984). *Abnormal psychology and modern life* (7th ed.). Glenview, Illinois: Scott.

Collaer, M.L. & Evans, J.R. (1982). A measure of short-term memory based on the WISC-R Coding subtest. *Journal of Clinical Psychology, 38*(3), 641-644.

Condell, J.F. (1963). The Bender Gestalt Test with mentally retarded children using the Koppitz Revised Scoring System. *Journal of Clinical Psychology, 19*, 430-431.

Cooper, J.A. & Sagar, H.J. (1993). Incidental and intentional recall in Parkinson's Disease: An account based on diminished attentional resources. *Journal of Clinical and Experimental Neuropsychology, 15*(5), 713-731.

Cooper, J.A., Sagar, H.J. & Sullivan, E.V. (1993). Short-term memory and temporal ordering in early Parkinson's disease: Effects of disease chronicity and medication. *Neuropsychologia*, *31*(9), 933-949.

Cooper, J.R. & Barnes, E.J. (1966). Technique for measuring reproductions of visual stimuli: II. Adult reproductions of the Bender-Gestalt. *Perceptual and Motor Skills*, *23*, 1135-1138.

Cooper, J.R., Dwarshuis, L. & Blechman, G. (1967). Technique for measuring reproductions of visual stimuli: III: Bender-Gestalt and severity of neurological deficit. *Perceptual and Motor Skills*, *25*, 506-508.

Cotler, S. & Palmer, R. J. (1970). The effects of test anxiety, sex of subject, and type of verbal reinforcement on maze performance of elementary school children. *Journal of Personality Assessment*, *38*, 216-234.

Crawford-Nutt, D.H. (1977). The effect of educational level on the test scores of people in South Africa. *Psychologia Africana*, *17*, 49-59.

Cronbach, L.J. (1991). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.

Culbertson, F.M. & Gunn, R.C. (1966). Comparison of the Bender Gestalt Test and Frostig Test in several clinical groups of children. *Journal of Clinical Psychology*, *22*, 439.

Davidson, L.A. (1974). Introduction. In R.M. Reitan & L.B. Davidson (Eds), *Clinical neuropsychology: Current status and applications* (pp.7-12). Washington: Winston.

Dana, R.H., Field, K. & Bolton, B. (1983). Variations of the Bender-Gestalt test: Implications for training in practice. *Journal of Personality Assessment*, *47*(1), 76-84.

Delaney, R.C. (1982). Screening for organicity: The problem of subtle neuropsychological deficit and diagnosis. *Journal of Clinical Psychology*, *38*(4), 843-846.

Dennis, M., Spiegler, B.J., Fitz, C.R., Hoffman, H.J., Hendrick, E.B., Humphreys, R.P. & Chuang, S. (1991). Brain tumours in children and adolescents-II. The neuroanatomy of deficits in working, associative and serial-order memory. *Neuropsychologia*, *29*(9), 829-847.

Diamond, R., White, R.F., Myers, R.H., Mastromauro, C., Koroshetz, W.J., Butters, N., Rothstein, D.M., Moss, M. & Vasterling, J. (1992). Evidence of presymptomatic cognitive decline in Huntington's disease. *Journal of Clinical and Experimental Neuropsychology, 14*(6), 961-975.

Dibner, A.S. & Korn, E.J. (1969). Group administration of the Bender Gestalt Test to predict early school performance. *Journal of Clinical Psychology, 25*, 263-268.

Dierks, D. & Cushna, B. (1969). Sex differences in the Bender Gestalt performance of children. *Perceptual and Motor Skills, 28*, 19-22.

Dinklage, D.K., Kantor, K. & Grodzinsky, G. (1992). Developmental aspects of the immediate recall of the Rey figure: Incidental versus intentional recall. *Journal of Clinical and Experimental Neuropsychology, 14*(1), 110.

Dodrill, C.B. (1979). Sex differences on the Halstead-Reitan neuropsychological battery and on other neuropsychological measures. *Journal of Clinical Psychology, 35*, 236-241.

Doubros, S.G. & Mascarenhas, J. (1969). Relations among Wechsler Full-Scale Scores, organicity-sensitive subtest scores and Bender-Gestalt error scores. *Perceptual and Motor Skills, 29*, 719-722.

Dyall, C.K. (1993). *The establishment of age stratified normative scores for children on the Bender Gestalt Test used as a measure of visual memory.* Unpublished Honours Project, Rhodes University, Grahamstown.

Edwards, S.L. (1980). The effects of stating expectations and varying the type of feedback during practice copying of Bender Gestalt designs on the performance, immediate and delayed recall, affective reaction, and self evaluation of performance of black first graders (Doctoral dissertation, Wayne State University, 1980). *Dissertation Abstracts International, 41B*(10), 3885-3886.

Ellis, A.W. & Young, A.W. (1988). *Human Cognitive Neuropsychology.* London: Lawrence Erlbaum Associates.

Eno, L. & Deichmann, J. (1980). A review of the Bender Gestalt Test as a screening instrument for brain damage with school-aged children of normal intelligence. *The Journal of Special Education, 14*(1), 37-45.

Ernst, J. (1987). Neuropsychological problem solving skills in the elderly. *Psychology and Aging, 2*, 363-365.

Even, A., Kipper, D.A. & Yehuda, S. (1988). Recall of figure "A" of the Bender-Gestalt Test among delinquents. *Journal of Clinical Psychology, 44*, 988-994.

Fiedler, M.F. & Schmidt, E.P. (1969). Sex differences in Bender-Gestalt drawings of seven-year-old children. *Perceptual and Motor Skills, 29*, 753-754.

Field, K., Bolton, B. & Dana, R.H. (1982). An evaluation of three Bender-Gestalt scoring systems as indicators of psychopathology. *Journal of Clinical Psychology, 38*(4), 838-842.

Finch Jnr, A.J., Anderson, J. & Kendall, P.C. (1976). Anxiety and Digit Span performance in emotionally disturbed children. *Journal of Consulting and Clinical Psychology, 44*(5), 874.

Finch, Jnr, A.J., Spirito, A., Garrison, S. & Marshall, P. (1983). Developmental differences in Bender-Gestalt recall of children with learning and behaviour problems. *Perceptual and Motor Skills, 56*, 87-90.

Finlayson, M.A.J., Johnson, K.A. & Reitan, R.M. (1977). Relationship of level of education to neuropsychological measures in brain-damaged and non-brain-damaged adults. *Journal of Consulting and Clinical Psychology, 45*(4), 536-542.

Fisher, C.B. (1993). Joining science and application: Ethical challenges for researchers and practitioners. *Professional Psychology: Research and Practice, 24*(3), 378-381.

Flick, G.L. & Duncan, C. (1973). Perceptual-motor dysfunction in children with sickle cell trait. *Perceptual and Motor Skills, 36*, 234.

Freed, E.X. (1966). Susceptibility of individual Bender-Gestalt test designs to rotation by psychiatric patients. *Journal of Clinical Psychology, 22*, 98-99.

Frey, P.D. & Pinelli Jnr, B. (1991). Visual discrimination and visuomotor integration among two classes of Brazilian children. *Perceptual and Motor Skills, 72*, 847-850.

Friedman, J., Strochak, R.D., Gitlin, S. & Gottsagen, M.L. (1967). Koppitz' Bender scoring system and brain injury in children. *Journal of Clinical Psychology, 23*, 179-182.

Friedt, L.R. & Gouvier, W.D. (1989). Bender Gestalt screening for brain dysfunction in a forensic population. *Criminal Justice and Behaviour, 16*(4), 455-464.

Frostig, M., Lefever, D.W. & Whittlesey, J.R.B. (1961). A developmental test of visual perception for evaluating normal and neurologically handicapped children. *Perceptual and Motor Skills, 12*, 383-394.

Furr, K.D. (1970). Standard scores for the Koppitz developmental scoring system. *Journal of Clinical Psychology, 26*, 78-79.

Gainotti, G. & Marra, C. (1994). Some aspects of memory disorders clearly distinguish dementia of the Alzheimer's type from depressive pseudo-dementia. *Journal of Clinical and Experimental Neuropsychology, 16*(1), 65-78.

Garron, D.C. & Cheifetz, D.I. (1965). Comment on "Bender Gestalt discernment of organic pathology". *Psychological Bulletin, 63*(3), 197-200.

Gass, C.S. & Russell, E.W. (1986). Differential impact of brain damage and depression on memory test performance. *Journal of Consulting and Clinical Psychology, 54*(2), 261-263.

Gilbert, J.G. & Levee, R.F. (1967). Performances of deaf and normally-hearing children on the Bender-Gestalt and the Archimedes Spiral Tests. *Perceptual and Motor Skills, 24*, 1059-1066.

Goff, A.F. & Parker, A.W. (1969). Reliability of the Koppitz scoring system for the Bender Gestalt Test. *Journal of Clinical Psychology, 25*, 407-409.

Goldberg, L.R. (1959). The effectiveness of clinicians judgments: The diagnosis of organic brain damage from the Bender-Gestalt Test. *Journal of Consulting Psychology, 23*(1), 25-33.

Golden, C.J. (1976). The identification of brain damage by an abbreviated form of the Halstead-Reitan neuropsychological battery. *Journal of Clinical Psychology, 32*(4), 821-826.

Goodstein, L.D., Spielberger, C.D., Williams, J.E. & Dahlstrom, W.G. (1955). The effects of serial position and design difficulty on recall of the Bender-Gestalt Test designs. *Journal of Consulting Psychology, 19*(3), 230-234.

Gowens, E. (Ed.) (1977). *Fowlers Modern English Usage* (2nd ed.). Oxford University Press.

Griffith, R.M. & Taylor, V.H. (1960). Incidence of Bender-Gestalt figure rotations. *Journal of Consulting Psychology, 24*(2), 189-190.

Gronwall, D. (1989). Behavioural assessment during the acute stages of traumatic brain injury. In M.D. Lezak (Ed.), *Assessment of the behavioural consequences of head trauma* (pp.19-36). New York: Liss.

Grossi, D., Correra, G., Calise, C., Ruscitto, M.A., Vecchione, V., Vigliardi, M.V. & Nolfe, G. (1993). Evaluation of the influence of illiteracy on neuropsychological performances by elderly persons. *Perceptual and Motor Skills, 77*, 859-866.

Haaland, K.Y., Temkin, N., Randahl, G. & Dikmen, S. (1994). Recovery of simple motor skills after head injury. *Journal of Clinical and Experimental Neuropsychology, 16*(3), 448-456.

Hain, J.D. (1964). A scoring method for identifying brain damage. *Journal of Consulting Psychology, 28*(1), 34-40.

Hargrove, D.S. & Breazeale, R.L. (1993). Psychologists and rural services: Addressing a new agenda. *Professional Psychology: Research and Practice, 24*(3), 319-324.

Harris, D.B. (1963). *Children's drawings as measures of intellectual maturity: A revision and extension of the Goodenough Draw-A-Man Test.* New York: Harcourt, Brace & World.

Hart, L.R. (1993). Misleading report of the poor screening power of the Bender Gestalt: Comment on Margolis, Williger, Greenlief, Dunn & Gfeller (1989). *Journal of Psychology, 128*(2), 185-187.

Hanvik, L.J. (1951). A note on rotations in the Bender Gestalt Test as predictors of EEG abnormalities in children. *Journal of Clinical Psychology, 7*, p.194.

Haynes, J.R. (1970). Factor-analytic study of performance on the Bender-Gestalt. *Journal of Consulting and Clinical Psychology, 34*, 345-347.

Heaton, R.K., Lyle, E., Baade, L.E. & Johnson, K.L. (1976). Neuropsychological test results associated with psychiatric disorders in adults. *Psychological Bulletin, 85*(1), 141-162.

Hebb, D.O. (1949). *The organization of behaviour: A neuropsychological theory.* New York: Wiley.

Henderson, N.B., Butler, B.V. & Goffeney, B. (1969). Effectiveness of the WISC and Bender-Gestalt Test in predicting arithmetic and reading achievement for white and nonwhite children. *Journal of Clinical Psychology, 25*, 268-271.

Hilgert, L.D. (1985). A Graphic Analysis Bender Gestalt Test. *Journal of Clinical Psychology, 41*(4), 505-511.

Holland, T.R. & Wadsworth, H.M. (1979). Comparison and combination of recall and Background Interference Procedures for the Bender-Gestalt Test with brain-damaged and schizophrenic patients. *Journal of Personality Assessment, 43*(2), 123-127.

Holroyd, J. (1966a). Cross validation of the Quast and Koppitz Bender-Gestalt signs of cerebral dysfunction. *Journal of Clinical Psychology, 22*, 200.

Holroyd, R.G. (1966b). The Koppitz Bender Gestalt Test for young children: A scoring guide. *Journal of Clinical Psychology, 22*, 440-441.

Houx, P.J. & Jolles, J. (1993). Age-related decline of psychomotor speed: Effects of age, brain health, sex and education. *Perceptual and Motor Skills*, 76, 195-211.

Huppert, F.A. & Beardsall, L. (1993). Prospective memory impairment as an early indicator of dementia. *Journal of Clinical and Experimental Neuropsychology, 15*(5), 805-821.

Hutt, M.L. & Briskin, G.J. (1960). *The clinical use of the Revised Bender-Gestalt Test.* New York: Grune & Stratton.

Hutton, J.B. (1966). Bender recall of children as related to age and intelligence. *Perceptual and Motor Skills, 23*, 34.

Imm, P.S., Foster, K.Y. Belter, R.W. & Finch Jnr, A.J. (1991). Assessment of short-term visual memory in child and adolescent psychiatric inpatients. *Journal of Clinical Psychology, 47*(3), 440-443.

Isaac, B.K. (1973). Perceptual-Motor development of first graders as related to class, race, intelligence, visual discrimination, and motivation. *Journal of School Psychology, 11*(1), 47-56.

Jambaqué, I., Dellatolas, G., Dulac, O., Ponsot, G. & Signoret, J.L. Verbal and visual memory impairment in children with epilepsy. (1993). *Neuropsychologia, 31*(12), 1321-1337.

Jernigan, A.J. (1967). Rotation style on the Bender Gestalt Test. *Journal of Clinical Psychology, 23,* 176-179.

Johnson, J.H., Mc Cusker, C. & Boyd, C. (1976). A validation study of the clinical screening examination for memory. *Journal of Clinical Psychology, 32,* 90-91.

Johnson, J.E., Hellkamp, D.T. & Lottman, T.J. (1971). The relationship between intelligence, brain damage, and Hutt-Briskin error on the Bender-Gestalt. *Journal of Clinical Psychology, 27,* 84-85.

Kampen, D.L & Grafman, J. (1989). Neuropsychological evaluation of penetrating head injury. In M.D. Lezak (Ed.), *Assessment of the behavioural consequences of head trauma* (pp.49-60). New York: Liss.

Kaplan, H.I. & Sadock, B.J. (1991). *Synopsis of psychiatry: Behavioral sciences clinical psychiatry.* Baltimore: Williams & Wilkins.

Kaspar, J.C. & Lampel, A.K. (1972). Interrater reliability for scoring the Bender Gestalt using the Koppitz method. *Perceptual and Motor Skills, 34,* 765-766.

Kay, T. & Silver, S.M. (1989). Closed head trauma: Assessment for rehabilitation. In M.D. Lezak (Ed.), *Assessment of the behavioural consequences of head trauma* (pp. 145-170). New York: Liss.

Kelly, K. & Van Vlaenderen, H. (February 1994). *Dynamics of participation in a community health project.* Paper presented at the Psychology and Transformation Conference, University of Western Cape.

Kempen, J.H. Kritchevsky, M. & Feldman, S.T. (1994). Effect of visual impairment on neuropsychological test performance. *Journal of Clinical and Experimental Neuropsychology, 16*(2), 223-231.

Kendall, B.S. (1962). Memory-for-Designs performance in the seventh and eighth decades of life. *Perceptual and Motor Skills, 14*, 399-405.

Kennedy, K.J. (1981). Age effects on Trail Making Test performance. *Perceptual and Motor Skills, 52*, 671-675.

Keogh, B.K. (1965). The Bender Gestalt as a predictive and diagnostic test of reading performance. *Journal of Consulting Psychology, 29*(1), 83-84.

Keogh, B.K. & Smith, C.E. (1961). Group techniques and proposed scoring system for the Bender-Gestalt Test with children. *Journal of Clinical Psychology, 17*, 172-175.

Keogh, B.K. & Smith, C.E. (1968). Changes in copying ability of young children. *Perceptual and Motor Skills, 26*, 773-774.

Köhler, W. (1920). Physical Gestalten (Die physischen Gestalten in Ruhe und im stationären Zustand, Eine naturphilsophische Untersuchung, Erlangen). In W.D. Ellis (Ed. and Trans.) (1938). *A source book of gestalt psychology* (pp.17-54). London: Kegan Paul.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology.* New York: The New American Library.

Kolb, B. & Whishaw, I.Q. (1990). In A.A. Boulton, G.B. Baker & M. Hiscock (Eds), *Neuromethods 17: Neuropsychology* (pp.1-36). New York: Freeman.

Koppitz, E.M. (1962). Diagnosing brain damage in young children with the Bender Gestalt Test. *Journal of Consulting Psychology, 26*(6), 541-546.

Koppitz, E.M. (1975). *The Bender Gestalt Test for young children. Vol II, research and application, 1963-1973.* New York: Grune and Stratton.

Koppitz, E.M. (1975). Bender Gestalt Test, Visual Aural Digit Span Test and reading achievement. *Journal of Learning Disabilities, 8*(3), 32-35.

Korin, H. (1974). Comparison of psychometric measures in psychiatric patients using heroin and other drugs. *Journal of Abnormal Psychology, 83*(2), 208-212.

Korman, M. & Blumberg, S. (1963). Comparative efficiency of some tests of cerebral damage. *Journal of Consulting Psychology, 27*(4), 303-309.

Kriegler, S. (1993). Options and directions for psychology within a framework for mental health services in South Africa. *South African Journal of Psychology, 23*(2), 61-70.

Lacks, P. (1982). Continued clinical popularity of the Bender-Gestalt Test: Response to Bigler and Ehrfurth. *Professional Psychology, 13*(5), 677-680.

Lacks, P. (1984). *Bender-Gestalt screening for brain dysfunction.* Missouri: Wiley.

Lacks, P.B., Colbert, J., Harrow, M. & Levine, J. (1970). Further evidence concerning the diagnostic accuracy of the Halstead Organic Test Battery. *Journal of Clinical Psychology, 26*, 480-481.

Lacks, P.B. & Newport, K. (1980). A comparison of scoring systems and level of scorer experience on the Bender-Gestalt Test. *Journal of Personality Assessment, 44*(4), 351-357.

Lacks, P. & Storandt, M. (1982). Bender Gestalt performance of normal older adults. *Clinical Psychology, 38*, 624-627.

Landis, B., Baxter, J. & Patterson Jnr, R.H. (1974). Bender Gestalt evaluation of brain dysfunction following open-heart surgery. *Journal of Personality Assessment, 38*(6), 556-562.

Larrabee, G.L., Youngjohn, J.R., Sudilovsky, A. & Crook III, T.H. (1993). Accelerated forgetting in Alzheimer-type dementia. *Journal of Clinical and Experimental Neuropsychology, 15*(5), 701-712.

Law, W.A., Martin, A., Mapou, R.L., Roller, T.L., Salazar, T.A.M., Temoshok, L.A. & Rundell, R. (1994). Working memory in individuals with HIV infection, *Journal of Clinical and Experimental Neuropsychology, 16*, 173-182.

Leckliter, I.N. & Matarazzo, J.D. (1989). The influence of age, education, IQ, gender and alcohol abuse on Halstead-Reitan Neuropsychological Test Battery performance. *Journal of Clinical Psychology, 45*(4), 484-512.

Levin, S.H., Eisenberg, H.M. & Miner, M.E. (1983). Neuropsychologic findings in head injured children. In K.S. Shapiro (Ed.), *Paediatric Head Trauma* (pp.223-241). New York: Futura.

Levine, J. & Feirstein, A. (1972). Differences in test performance between brain-damaged, schizophrenic, and medical patients. *Journal of Consulting and Clinical Psychology, 39*(3), 508-511.

Leviton, I., Kirby, C., Guild-Wilson, M. & Neff, R.K. (1993). The Boston Teacher Questionnaire 2. Assessments of Validity. *Journal of Child Neurology, 8*, 54-63.

Lezak, M.D. (1983). *Neuropsychological assessment* (2nd ed.). Oxford University Press.

Lezak, M.D. (1989a). Introduction: The relevance of behavioural assessment in head injury. In M.D. Lezak (Ed.), *Assessment of the behavioural consequences of head trauma* (pp. xi-xiii). New York: Liss.

Lezak, M.D. (1989b). Assessment of psychosocial dysfunctions resulting from head trauma. In M.D. Lezak (Ed.), *Assessment of the behavioural consequences of head trauma* (pp. 113-143). New York: Liss.

Liddell, G. & Kvalsvig, J. (1990). Science and social accountability: Issues related to South African developmental psychology. *South African Journal of Psychology, 29*, 1-9.

Lishman, W.A. (1978). *Organic psychiatry: The psychological consequences of cerebral disorder.* Oxford: Blackwell Scientific Publications.

Loewenstein, D.A., Argüelles, X.T., Argüelles, S. & Linn-Fuentes, P. (1994). Potential cultural bias in the neuropsychological assessment of the older adult. *Journal of Clinical and Experimental Neuropsychology, 16*(4), 623-629.

Lubin, B., Larsen, R.M. & Matarazzo, J.D. (1984). Patterns of psychological test usage in the United States 1935-1982. *American Psychologist*, April, 451-454.

Lubin, B., Larsen, R.M., Matarazzo, J.D. & Seever, M. (1985). Psychological test usage patterns in five professional settings. *American Psychologist*, 857-861.

Lubin, B. & Sands, E.W. (1992). Bibliography of the psychometric properties of the Bender Visual-Motor Gestalt Test 1970, 1991. *Perceptual and Motor Skills, 75*, 386-386.

Lyle, O.E. & Gottesman, I.I. (1977). Premorbid psychometric indicators of the gene for Huntington's disease. *Journal of Consulting and Clinical Psychology, 45*(6), 1011-1022.

Lyle, O.E. & Quast, W. (1976). The Bender Gestalt: Use of clinical judgment versus recall scores in prediction of Huntington's disease. *Journal of Consulting and Clinical Psychology, 44*(2), 229-232.

Maloney, M.P. & Ward, E. (1970). Bender-Gestalt Test performance of 'organic' and 'functional' mentally retarded subjects. *Perceptual and Motor Skills, 31*, 860.

Mapou, R.L. & Law, W.A. (1994). Neurobehavioural aspects of HIV disease and AIDS: An update. *Professional Psychology: Research and Practice, 25*(2), 132-140.

Margolis, R.B., Williger, N.R., Greenlief, C.L., Dunn, E.J. & Gfeller, J.D. (1989). The sensitivity of the Bender-Gestalt Test as a screening instrument for neuropsychological impairment in older adults. *Journal of Psychology, 123*(2), 179-186.

Markowitsch, H.J., Calabrese, P., Haupts, M., Durwen, H.F., Liess, J. & Gehlen, W. (1993). Searching for the anatomical basis of retrograde amnesia. *Journal of Clinical and Experimental Neuropsychology, 15*(6), 947-967.

Marsh, G.G. (1972). Impaired visual-motor ability of children with Duchenne muscular dystrophy. *Perceptual and Motor Skills, 35*, 504-506.

Marsico, D.S. & Wagner, E.E. (1990). A comparison of the Lacks and Pascal-Suttell Bender-Gestalt scoring methods for diagnosing brain damage in an outpatient sample. *Journal of Clinical Psychology, 46*(6), 868-877.

Marteau, T.M. (1994). Psychology and screening; Narrowing the gap between efficacy and effectiveness. *British Journal of Clinical Psychology, 32*, 1-10.

Mc Kay, M.F. & Neale, M.D. (1985). Predicting early school achievement in reading and handwriting using major 'error' categories from the Bender-Gestalt Test for young children. *Perceptual and Motor Skills, 60*, 647-654.

Mc Connell, O.L. (1967). Koppitz' Bender-Gestalt scores in relation to organic and emotional problems in children. *Journal of Clinical Psychology, 23*, 370-374.

McCarthy, D.P. (1975). The feasibility of a group Bender-Gestalt Test for preschool and primary school-aged children. *Journal of School Psychology, 13*(2), 134-141.

McFie, J. (1975). *Assessment of organic intellectual impairment.* London: Academic Press.

McIntosh-Michaelis, S.A., Roberts, M.H., Wilkinson, S.M., Diamond, I.D., McLellan, D.L., Martin, J.P. & Spackman, A.J. (1991). The prevalence of cognitive impairment in a community survey of multiple sclerosis. *British Journal of Clinical Psychology, 30*, 333-348.

Mc Manis D.L. (1974). Memory-for-Designs performance of brain-damaged and non-damaged psychiatric patients. *Perceptual and Motor Skills, 38*, 847-852.

Mc Manis, D.L., Figley, C., Richert, M. & Fabre, T. (1978). Memory-for-Designs, Bender Gestalt, Trail Making Test and WISC-R performance of retarded and adequate readers. *Perceptual and Motor Skills, 46*, 443-450.

McNamara, J.R., Porterfield, C.L. & Miller, L.E. (1969). The relationship of the Wechsler Preschool and Primary Scale of Intelligence with the Coloured Progressive Matrices (1956) and the Bender Gestalt Test. *Journal of Clinical Psychology, 25*, 65-68.

Meehl, P.E. & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*(3), 194-216.

Meltzer, M.L. (1983). Poor memory: A case report. *Journal of Clinical Psychology, 39*(1), 3-10.

Miller, L.L. & Branconnier, R.J. (1983). Cannabis: Effects on memory and the cholinergic limbic system. *Psychological Bulletin, 93*(3), 441-456.

Miller, L.J. & Hutt, M.L. (1975). Psychopathology Scale of the Hutt adaptation of the Bender-Gestalt Test: Reliability. *Journal of Personality Assessment, 39*(2), 129-131.

Miller, L.C., Loewenfeld, R., Linder, R. & Turner, J. (1963). Reliability of Koppitz' scoring system for the Bender Gestalt. *Journal of Clinical Psychology, 19*, 211.

Morsbach, G., Del Priori, C. & Furnell, J. (1975). Two aspects of scorer reliability in the Bender-Gestalt Test. *Journal of Clinical Psychology, 31*, 90-93.

Mosher, D.I. & Smith, J.P. (1965). The usefulness of two scoring systems for the Bender Gestalt Test for identifying brain damage. *Journal of Consulting Psychology, 29*(6), 530-536.

Naugle, R.I., Cullum, C.M. & Bigler, E.D. (1990). Evaluation of intellectual and memory function among dementia patients who were intellectually superior. *The Clinical Neuropsychologist, 4*(4), 355-374.

Nell, V. (1994). Interpretation and misinterpretation of the South African Wechsler-Bellevue Adult Intelligence Scale: A history and a prospectus. *South African Journal of Psychology, 24*(2), 100-109.

Nielson, S. & Sapp, G.L. (1991). Bender-Gestalt developmental scores: Predicting reading and mathematics achievement. *Psychological Reports, 69*, 39-42.

Oas, P. (1984). Validity of the Draw-a-Person and Bender Gestalt Tests as measures of impulsivity with adolescents. *Journal of Consulting and Clinical Psychology, 52*(6), 1011-1019.

Obrzut, J.E., Taylor, H.D. & Thweatt, R.C. (1972). Re-examination of Koppitz' Developmental Bender Scoring System. *Perceptual and Motor Skills, 34*, 279-282.

Oddy, M. (1984). Head injury during childhood: The psychological implications. In N. Brooks (Ed.), *Closed head injury: Psychological, social and family consequences* (pp.179-194). Oxford University Press.

Olin, T.D. & Reznikoff, M. (1957). Quantification of the Bender-Gestalt Recall: A pilot study. *Journal of Projective Techniques, 21*, 265-277.

Oscar-Berman, M., Clancy, J.P. & Weber, D.A. (1993). Discrepancies between IQ and memory scores in alcoholism and aging. *The Clinical Neuropsychologist, 7*(3), 281-296.

Pang, D. (1989). Physics and pathophysiology of closed head injury. In M.D. Lezak (Ed.), *Assessment of the behavioural consequences of head trauma* (pp.1-17). New York: Liss.

Parsons, L. & Weinberg, S.L. (1993). The Sugar scoring system for the Bender-Gestalt test: An objective approach that reflects clinical judgement. *Perceptual and Motor Skills, 77*, 883-893.

Parsons, O.A. & Prigatano, G.P. (1978). Methodological considerations in clinical neuropsychological research. *Journal of Consulting and Clinical Psychology, 46*(4), 608-619.

Pascal, G.R. & Suttell, B.J. (1951). *The Bender Gestalt Test.* New York: Grune & Stratton.

Patton, J.E. & Offenbach, S.I. (1978). Effects of visual and auditory distracters on learning disabled and normal children's recognition memory performance. *Journal of Educational Psychology, 70*(5), 788-795.

Pauker, J.D. (1976). A quick-scoring system for the Bender-Gestalt: Interrater reliability and scoring validity. *Journal of Clinical Psychology, 32*, 86-89.

Peek, R.M. & Olson, G.W. (1955). The Bender-Gestalt Recall as an index of intellectual functioning. *Journal of Clinical Psychology, 11*, 185-188.

Petermann, B. (1932). *The Gestalt theory and the problem of configuration.* London: Kegan.

Ponsford, J. & Kinsella, G. (1992). Attentional deficits following closed-head injury. *Journal of Clinical and Experimental Neuropsychology, 14*(5), 822-838.

Pope, P. & Snyder, R.T. (1970). Modification of selected Bender designs and interpretation of the first graders' visual-perceptual maturation with implications for gestalt theory. *Perceptual and Motor Skills, 30*, 263-267.

Popplestone, J.A. (1956). Variability of the Bender Gestalt designs. *Perceptual and Motor Skills, 6*, 269-271.

Pribram, K.H. (1986). Nonlocality and localization in the primate forebrain. In S.B. Filskov & T.J. Boll (Eds), *Handbook of Clinical Neuropsychology* (pp.606-651). New York: Wiley.

Prigatano, G.P. & Parsons, O.A. (1976). Relationship of age and education to Halstead Test performance in different patient populations. *Journal of Consulting and Clinical Psychology, 44*(4), 527-533.

Purisch, A.D., Golden, C. J. & Hammeke, T.A. (1978). Discrimination of schizophrenic and brain-injured patients by a standardized version of Luria's Neuropsychological Tests. *Journal of Consulting and Clinical Psychology, 46*(6), 1266-1273.

Quast, W. (1961). The Bender Gestalt: A clinical study of children's records. *Journal of Consulting Psychology, 25*, 405-408.

Racusin, G.R. & Moss, N.E. (1991). Psychological assessment of children and adolescents. In M Lewis (Ed.), *Child and adolescent psychiatry: Comprehensive text book* (pp.472-485). Baltimore: Williams & Wilkins.

Rahmann, H. & Rahmann, M. (1991). *The neurobiological basis of memory and behaviour.* (translated from German by S.J. Freeman). New York: Springer.

Rector, M.T. (1984). The study of rotations and spatial displacements of certain Bender Gestalt figures by black children (Doctoral dissertation, New York University, 1984). *Dissertation Abstracts International, 45B* (6), 1935.

Reeves, D. & Wedding, D. (1994). The clinical assessment of memory. New York: Springer.

Reynolds, W.M. (1979). Psychological tests: Clinical usage versus psychometric quality. *Professional Psychology, 10*, 324-329.

Reznikoff, M. & Olin, T.D. (1957). Recall of the Bender-Gestalt designs by organic and schizophrenic patients: A comparative study. *Journal of Clinical Psychology, 13*, 183-187.

Rimm, D. (1963). Cost efficiency and test prediction. *Journal of Consulting Psychology, 27*(1), 89-91.

Ris, M.D. & Noll, R.B. (1994). Long-term neurobehavioural outcome in paediatric brain tumour patients: Review and methodological critique. *Journal of Clinical and Experimental Neuropsychology, 16*(1), 21-42.

Rogers, D.L. (1980). Bender Test recall in children: An unreliable test. *Perceptual and Motor Skills, 50*, 859-862.

Rogers, D.L. & Swenson, W.M. (1975). Bender-Gestalt recall as a measure of memory versus distractibility. *Perceptual and Motor Skills, 40*, 919-922.

Rosecrans, C.J. & Schaffer, H.B. (1969). Bender-Gestalt time and score differences between matched groups of hospitalized psychiatric and brain-damaged patients. *Journal of Clinical Psychology, 25*, 409-410.

Russell, E.W. (1976). The Bender-Gestalt and the Halstead-Reitan Battery:a case study. *Journal of Clinical Psychology, 32*(2), 355-361.

Ryckman, D.B., Rentfrow, R., Fargo, G. & Mc Cartin, R. (1972). Reliabilities of three tests of form-copying. *Perceptual and Motor Skills, 34*, 917-918.

Sabatino, D.A. & Becker, J.T. (1969). Relations among five tests of behavior. *Perceptual and Motor Skills, 29*, 487-490.

Sabatino, D.A. & Ysseldyke, J.E. (1972). Effect of extraneous 'background' on visual-perceptual performance of readers and nonreaders. *Perceptual and Motor Skills, 35*, 323-328.

Satz, P., Fennell, E. & Reilly, C. (1970). Predictive validity of six neurodiagnostic tests: A decision theory analysis. *Journal of Consulting and Clinical Psychology, 34*(3), 375-381.

Satz, P. & Fletcher, J.M. (1981). Emergent trends in neuropsychology: An overview. *Journal of Consulting and Clinical Psychology, 49*(6), 851-865.

Schachter, S., Brannigan, G.G. & Tooke, W. (1991). Comparison of two scoring systems for the modified version of the Bender-Gestalt Test. *Journal of School Psychology, 29*, 265-269.

Schilder, P. (1938). Preface. In L. Bender, *A Visual Motor Gestalt Test and its clinical use* (pp.vii-iv). New York: American Orthopsychiatric Association.

Schneider, W.E. & Pressley, M. (1988). *Memory development between 2 and 20.* New York: Springer.

Schulberg, H.C. & Tolor, A. (1961). The use of the Bender-Gestalt in clinical practice. *Journal of Projective Techniques, 25*, 347-351.

Sczechowicz, E. & Hinrichsen, J.J. (1980). Effects of instruction set on the Bender recall performance of learning disabled and normal children. *Journal of Personality Assessment, 44*(5), 465-468.

Segalowitz, S.J., Unsal, A. & Dywan, J. (1992). CNV evidence for the distinctiveness of frontal and posterior neural processes in a traumatic brain-injured population. *Journal of Clinical and Experimental Neuropsychology, 14*(4), 545-565.

Shaw, M.C. & Cruickshank, W.M. (1956). The use of the Bender Gestalt Test with epileptic children. *Journal of Clinical Psychology, 12*, 192-193.

Shearn, C.R., Berry, D.F. & Fitzgibbons, D.J. (1974). Usefulness of the Memory-for-Designs Test in assessing mild organic complications in psychiatric patients. *Perceptual and Motor Skills, 38*, 1099-1104.

Shein, J.H. (1975). The concurrent validity of Bender-Gestalt recall for differential assessment of organicity in children (Doctoral dissertation, Fordham University, 1975). *Dissertation Abstracts International, 36A*(8), 5162.

Shochet, I.M. (1994). The moderator effect of cognitive modifiability on a traditional undergraduate admissions test for disadvantaged black students in South Africa. *South African Journal of Psychology, 24*(4), 208-215.

Simensen, R.J. (1974). Correlations among Bender-Gestalt, WISC Block Design, Memory-for-Designs,and the Pupil Rating Scale. *Perceptual and Motor Skills, 38*, 1249-1250.

Simon, E., Leach, L., Winocur, G. & Moscovitch, M. (1994). Intact primary memory in mild to moderate Alzheimer Disease: Indices from the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology, 16*(3), 414-422.

Skeen, J.A., Strong, V.N. & Book, R.M. (1982). Comparison of learning disabled children's performance on Bender Visual-Motor Gestalt Test and Beery's Developmental Test of Visual Motor Integration. *Perceptual and Motor Skills, 55*, 1257-1258.

Smith, D.C. & Martin, R.A. (1967). Use of learning cues with the Bender Visual Motor Gestalt Test in screening children for neurological impairment. *Journal of Consulting Psychology, 31*(2), 205-209.

Smith, J.D. & Winnick, R.H. (1979). Practice effects within a multiple-phase Bender administration. *Perceptual and Motor Skills, 48*, 439-442.

Snortum, J.R. (1965). Performance of different diagnostic groups on the tachistoscopic and copy phases of the Bender-Gestalt. *Journal of Consulting Psychology, 29*(4), 345-351.

Snow, W.G. & Sheese, S. (1985). Lateralized brain damage, intelligence, and memory: A failure to find sex differences. *Journal of Consulting and Clinical Psychology, 53*(6), 940-941.

Snyder, R.T. & Kalil, J. (1968). Item analysis, inter-examiner reliability and scoring problems for Koppitz scoring on the Bender Gestalt for six-year-olds. *Perceptual and Motor Skills, 27*, 1351-1358.

Song, A.Y. & Song, R.H. (1969). The Bender-Gestalt Test with the Background Interference Procedure on mental retardates. *Journal of Clinical Psychology, 25*, 69-71.

Strain, G.S. & Kinzie, W.B. (1969). Reducing misdiagnosis of schizophrenic patients on a test for brain damage. *Journal of Clinical Psychology, 25*, 262.

Steinberg, O.A. (1990). *World Literacy Year 1990*. Johannesburg: English Literacy Year 1990.

Swensen, C.H. & Pascal, G.R. (1951). A note on the Bender-Gestalt Test as a prognostic indicator in mental illness. *Journal of Clinical Psychology, 9*, 193.

Swiercinsky, D.P. (1979). Factorial pattern description and comparison of functional abilities in neuropsychological assessment. *Perceptual and Motor Skills, 48*, 231-241.

Taylor, H.D. & Thweatt, R.C. (1972). Cross-cultural developmental performance of Navajo children on the Bender-Gestalt Test. *Perceptual and Motor Skills, 35*, 307-309.

Taylor, J.B. (1965). The Bender Gestalt as a measure of intelligence and adjustment in the lower intellectual range. *Journal of Consulting Psychology, 29*(6), 595.

Taylor, R.L., Kauffman, D. & Partenio, I. (1984). The Koppitz Developmental Scoring System for the Bender-Gestalt: Is it developmental? *Psychology in the Schools, 21*, 425-428.

Thomas, A.D. (1984). Bender scores and the Horse as a distinct item on Object Assembly of the WISC. *Perceptual and Motor Skills, 59*, 103-106.

Thweatt, R.C. (1963). Prediction of school learning disabilities through the use of the Bender Gestalt Test: A validation study of Koppitz's scoring technique. *Journal of Clinical Psychology, 19*, 216-217.

Tolor, A. (1956). A comparison of the Bender-Gestalt Test and the Digit-Span Test as measures of recall. *Journal of Consulting Psychology, 20*(4), 305-309.

Tolor, A. (1958). Further studies on the Bender-Gestalt Test and the Digit-Span Test as measures of recall. *Journal of Clinical Psychology, 14*, 14-18.

Tolor, A. & Brannigan, G.G. (1980). *Research and clinical applications of the Bender Gestalt Test.* Springfield: Charles Thomas.

Tröster, A.I., Butters, N., Salmon, D.P., Cullum, C.M., Jacobs, D., Brandt, J. & White, R.F. (1993). The diagnostic utility of savings scores: Differentiating Alzheimer's and Huntington's disease with the Logical Memory and Visual Reproduction tests. *Journal of Clinical and Experimental Neuropsychology, 15*(5), 773-788.

Vakil, E., Arbell, N., Gozlan, M., Hoofien, D. & Blachstein, H. (1992). Relative importance of informational units and their role in long-term recall by closed-head-injured patients and control groups. *Journal of Consulting and Clinical Psychology, 60*(5), 802-803.

Vane, J.R. (1967). An evaluation of the Harris revision of the Goodenough Draw-A-Man Test. *Journal of Clinical Psychology, 23*(3), 375-377.

Viljoen, G., Levett, A., Tredoux, C. & Anderson, S. (1994). Using the Bender Gestalt in South Africa: Some normative data for Zulu-speaking children. *South African Journal of Psychology, 24*(3), 145-151.

Vilkki, J., Ahola, K., Holst, P., Öhman, J., Servo, A. & Heiskanen, O. (1994). Prediction of psychosocial recovery after head injury with cognitive tests and neurobehavioural ratings. *Journal of Clinical and Experimental Neuropsychology, 16*(3), 325-338.

Vilkki, J., Holst, P., Öhman, J., Servo, A. & Heiskanen, O. (1992). Cognitive test performances related to early and late computed tomography findings after closed head injury. *Journal of Clinical and Experimental Neuropsychology, 14*(4), 518-532.

Wade, T.C. & Baker, T.B. (1977). Opinions and use of psychological tests. A survey of clinical psychologists. *American Psychologist, 32*, 874-882.

Wade, T.C., Baker, T.B., Morton, T.L. & Baker, L.J. (1978). The status of psychological testing in clinical psychology: Relationships between test use and professional activities and orientations. *Journal of Personality Assessment, 42*(1), 3-10.

Wagner, E.E. & Flamos, O. (1988). Optimized split-half reliability for the Bender Visual Motor Gestalt Test: Further evidence for the use of maximum procedure. *Journal of Personality Assessment, 52*(3), 454-458.

Wagner, E.E. & Marsico, D.S. (1991). Redundancy in the Pascal-Suttell Bender-Gestalt scoring system: Discriminating organicity with only one design. *Journal of Clinical Psychology, 47*(2), 261-263.

Wagner, E.E. & Murray, A.Y. (1969). Bender-Gestalts of organic children: Accuracy of clinical judgement. *Journal of Projective Techniques and Personality Assessment, 33*(3), 240-242.

Walsh, K. (1987). *Neuropsychology: A clinical approach* (2nd ed.). New York: Churchill Livingston.

Walsh, K. (1991). *Understanding brain damage: A primer of neuropsychological evaluation* (2nd ed.). New York: Churchill Livingston.

Weinstein, S. & Johnson, L. (1964). The Bender-Gestalt Test in differential diagnosis of temporal lobectomy and schizophrenia. *Perceptual and Motor Skills, 18*, 813-820.

Weiss, A.A. (1970). Reproduction from memory and frequency of recall of Bender-Gestalt figures in non-clinical subjects of different ages. *Israeli Annals of Psychiatry and Related Disciplines, 8*(2), 143-145.

Weiss, A.A. (1971a). Incidence of rotations of Bender-Gestalt figures in three age groups of normal Israeli school children. *Perceptual and Motor Skills, 32*, 691-694.

Weiss, A.A. (1971b). Directionality in four Bender-Gestalt figures: II. *Perceptual and Motor Skills, 32*, 412-414.

Welcher, D.W., Wessel, K.W., Mellits, E.D. & Hardy, J.B. (1974). The Bender-Gestalt Test as an indicator of neurological impairment in young inner-city children. *Perceptual and Motor Skills, 38*, 899-910.

Wertheimer, M. (1922). The general theoretical situation (Untersuchungen zur Lehre von der Gestalt). In W.D. Ellis (Ed. and Trans.) (1938), *A source book of gestalt psychology* (pp.12-16). London: Kegan Paul.

Wertheimer, M. (1923). Laws of organization in perceptual forms (Untersuchungen zur Lehre von der Gestalt). In W.D. Ellis (Ed. and Trans.) (1938), *A source book of gestalt psychology* (pp.71-88). London: Kegan Paul.

Wertheimer, M. (1925). Gestalt theory (Über Gestalttheorie). In W.D. Ellis (Ed. and Trans.) (1938), *A source book of gestalt psychology* (pp.1-11). London: Kegan Paul.

Wilkinson, C.A. & Koestler, R. (1983). Repeated recall: A new model and tests of its generality from childhood to old age. *Journal of Experimental Psychology: General*, *112*, 423-451.

Williams, D.T. (1991). Neuropsychiatric signs, symptoms and syndromes. In M. Lewis (Ed.), *Child and adolescent psychiatry: Comprehensive text book* (pp.340-346). Baltimore: Williams & Wilkins.

Williams, M. (1979). *Brain damage,behaviour and the mind*. Chichester: Wiley.

Wulf, F. (1922). *Tendencies in figured variation* (Über die Verändering van Vorstellungen (Gedächtnis und Gestalt). In W.D. Ellis (Ed. and Trans. (1930), *A source book of gestalt psychology* (pp.136-148). London: Kegan Paul.

Yeates, K.O. & Mortesen, M.E. (1994). Acute and chronic neuropsychological consequences of mercury vapor poisoning in early adolescents. *Journal of Clinical and Experimental Neuropsychology*, *16*(2), 209-222.

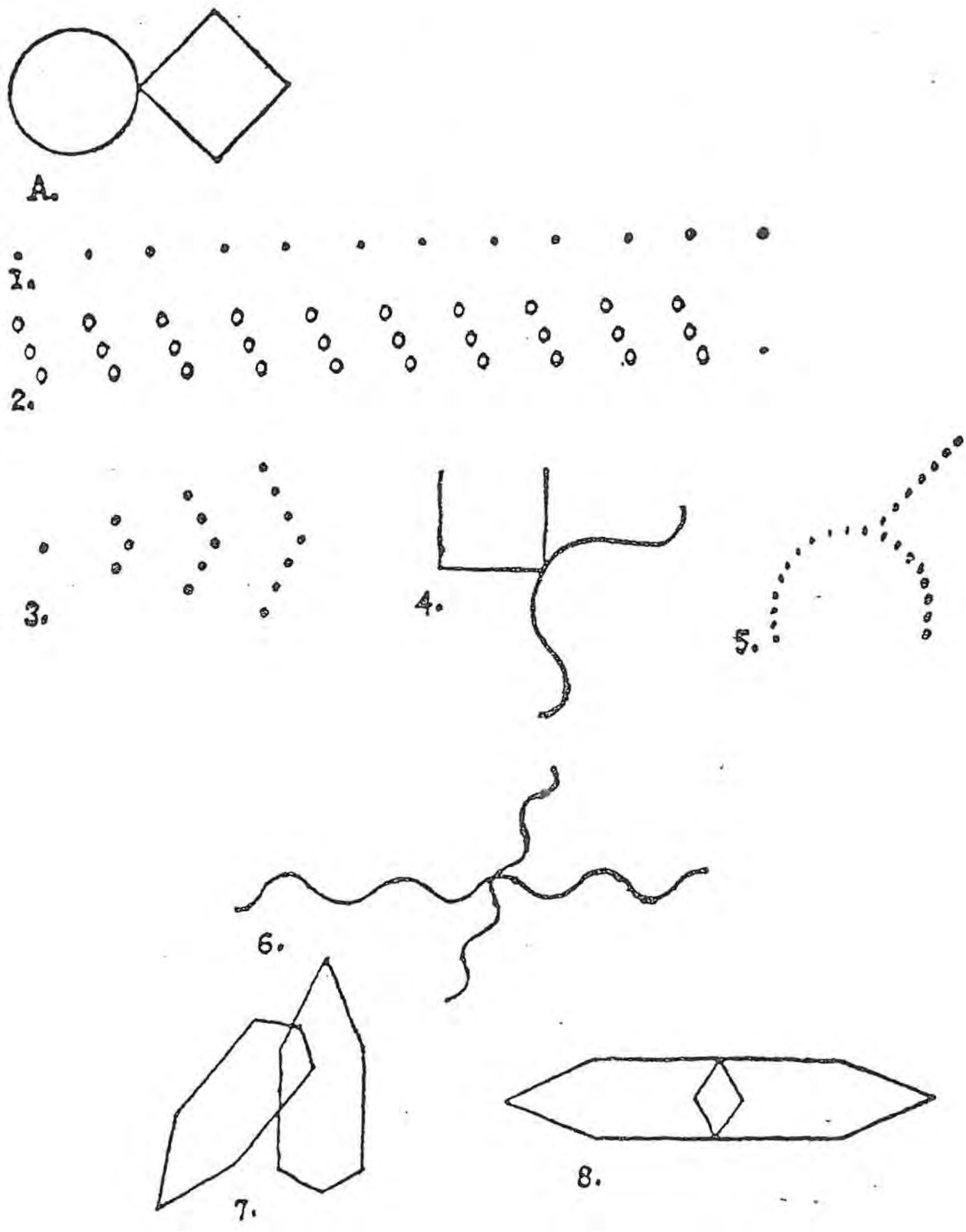
Youngjohn, J.R. & Crook III, T.H. (1993). Learning, forgetting and retrieval of everyday material across the adult life span. *Journal of Clinical and Experimental Neuropsychology*, *15*(4), 447-460.

Yousefi, B.F., Shahim, S., Razavieh, A., Mehryar A.H., Hosseini, A.A. & Alborzi, S. (1992). Some normative data on the Bender Gestalt Test performance of Iranian children. *British Journal of Educational Psychology*, *62*, 410-416.

Yulis, S. (1970). Performance of normal and organic brain-damaged subjects on the Canter Background Interference Procedure Test as a function of drive. *Journal of Consulting and Clinical Psychology*, *34*(2), 184-188.

Zuelzer, M.B., Stedman, J.M. & Adams, R. (1976). Koppitz Bender Gestalt scores in first-grade children as related to ethnocultural background, socioeconomic class, and sex factors. *Journal of Consulting and Clinical Psychology*, *44*(5), 875.

APPENDIX 1: THE BG DESIGNS



APPENDIX 2: LACKS'S (1984) SCORING SYSTEM

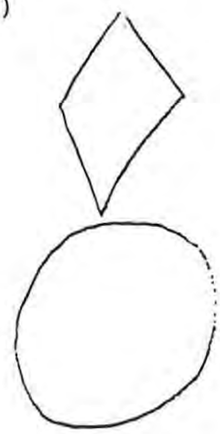
Lacks's (1984) scoring system from her book, *Bender Gestalt Screening for Brain Dysfunction*, Copyright © (1984, John Wiley & Sons, Inc.) is reproduced with the kind permission of the publisher (see Appendix 9B, p.265).

1. ROTATION ***Definition.*** Score if there is a rotation of 80° to 180° (including mirror-imaging) of the major axis of the whole figure (not a part of the figure). Do not score if *S* shifts the position of the card or the paper and then draws the figure accurately.

Figures. Score for all figures.

90° rotations

(A)

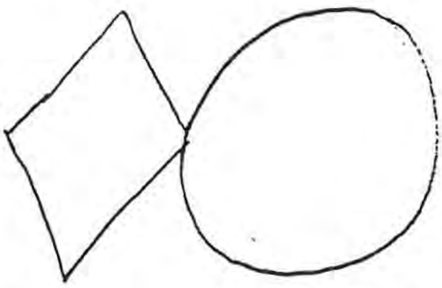


(1)

(3)

180° rotations

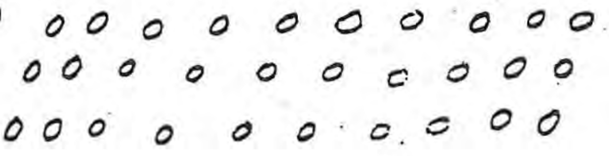
(A)



(3)

mirror-imaging

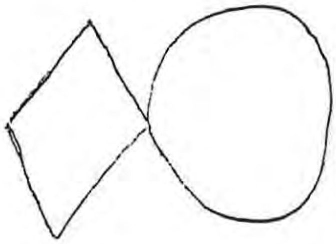
(2)



(5)

Do Not Score

(A)



(card was turned this way)

(3)

(approximately 45° rotation)

(5)

(only part of figure is rotated)

2. OVERLAPPING DIFFICULTY

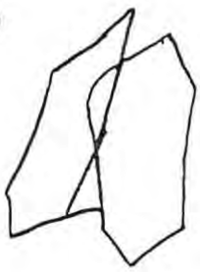
Definition. Difficulty in reproducing the portions of the figures that should overlap.

Figures. Score only for figures 6 and 7.

(7)

Omission of the portions of the figure which overlap

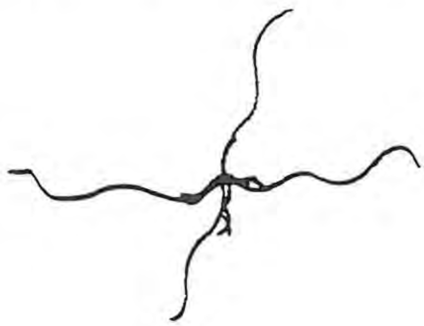
(7)



Simplification of figures only at the point of overlap

Marked sketching or reworking only at the point of overlap

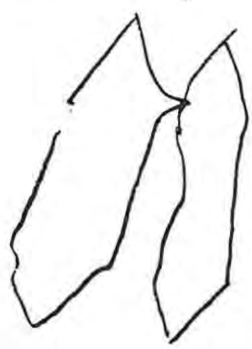
(6)

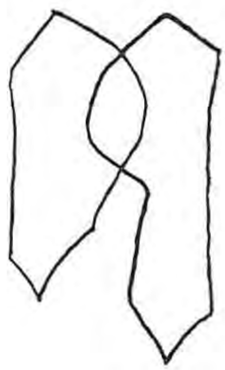


(7)

Distortion of the figure at the point of overlap

(7)

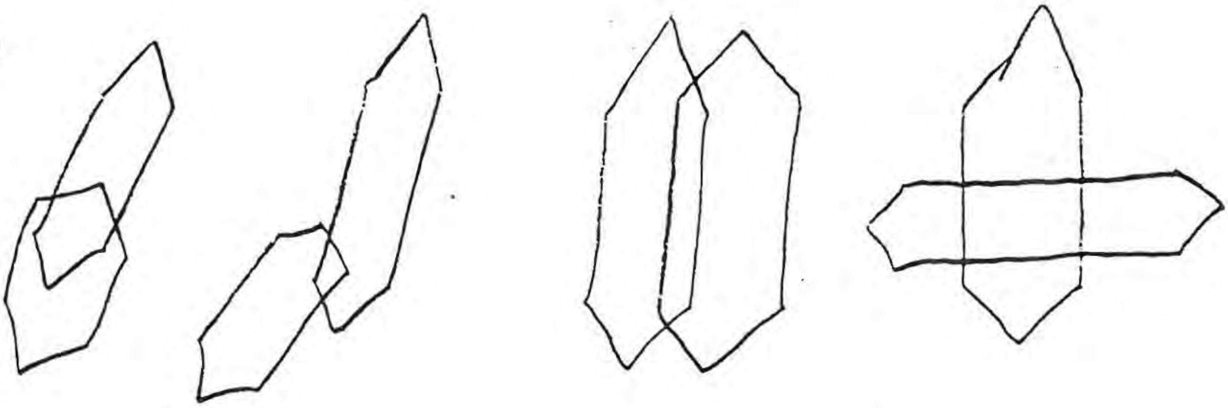




Figures overlap at the wrong place

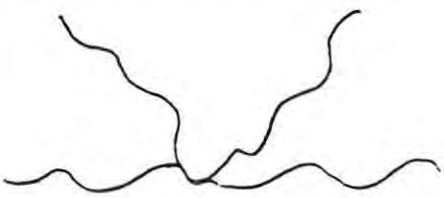
(6)

(7)

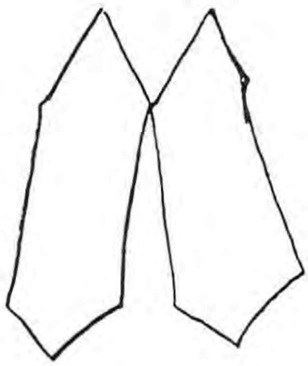


Failure of figures to overlap

(6)

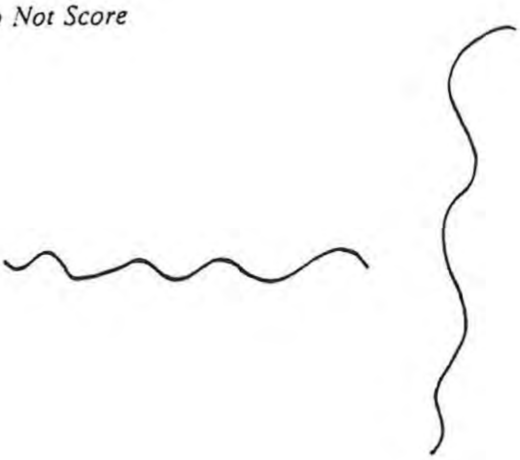


(7)

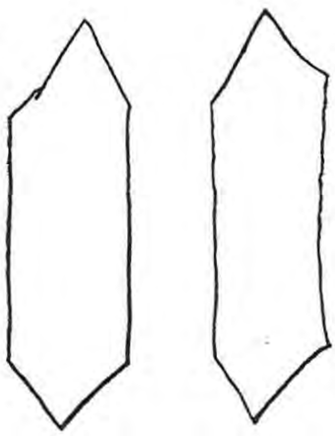


Do Not Score

(6)



(7)

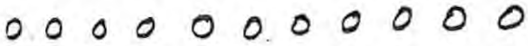


(parts of figures more than ⅛ inch apart, score Simplification)

3. SIMPLIFICATION ***Definition.*** Score if the figure is drawn in a simplified or easier form that is not more primitive, from a maturational point of view, from the stimulus.

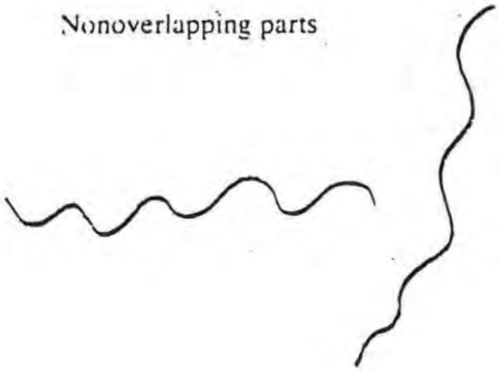
Figures. Score for all figures.

Circles for dots on figure 1

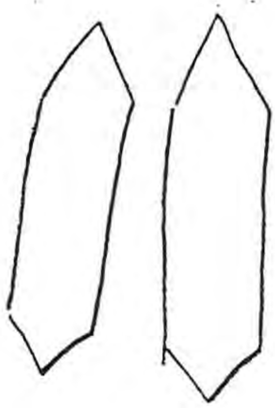


Nonoverlapping parts

(6)

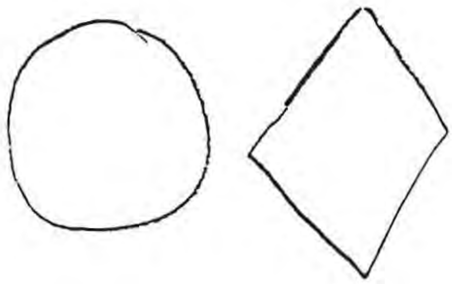


(7)

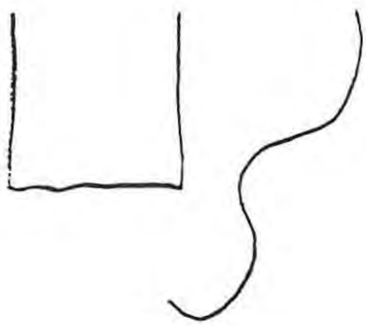


Joining parts of figures are more than 1/8 inch apart

(A)

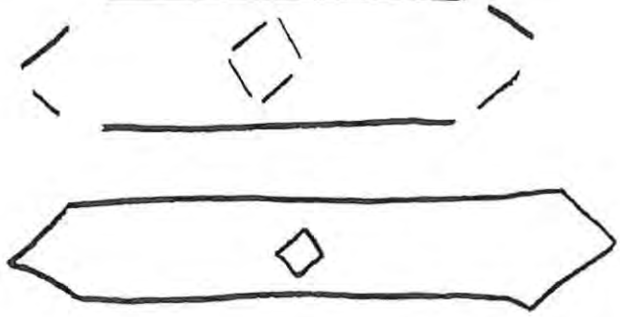


(4)



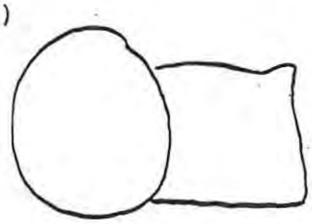
(5)

(8)



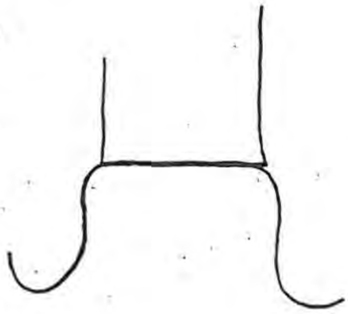
Very simplified drawing

(A)

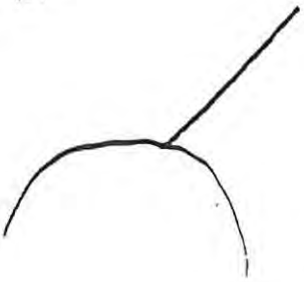


(3)

(4)

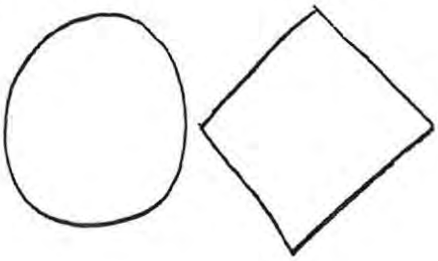


(5)



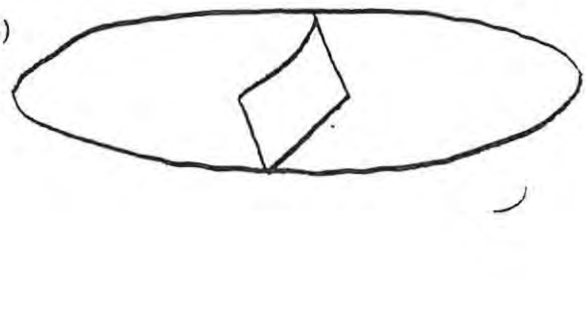
Do Not Score

(A)



(figures are less than 1/8 inch apart. score Closure Difficulty)

(8)



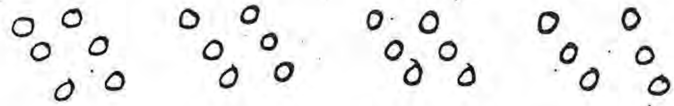
(curves substituted for angles. not an error)

4. FRAGMENTATION ***Definition.*** Score if the figure is broken up into parts destroying the gestalt or if the figure is incomplete (unless *S* refuses to draw the entire figure).

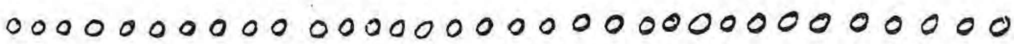
Figures. Score for all figures.

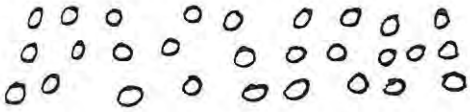
Figure broken into parts resulting in destruction of the gestalt

(1)


(2)


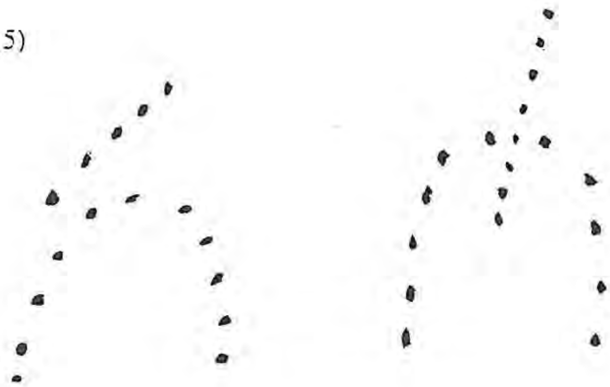
(one long row of 33 circles)

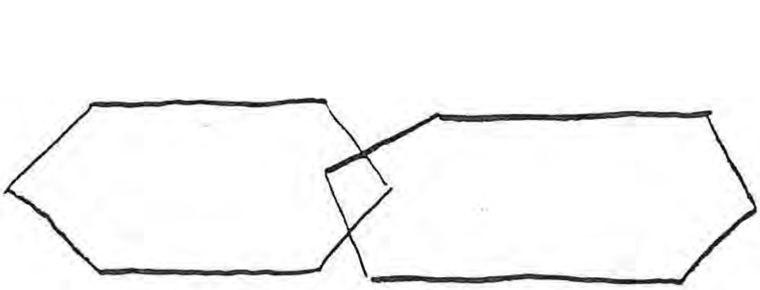
(2)


(2)


(random drawing of 33 circles with no recognition of the stimulus pattern)

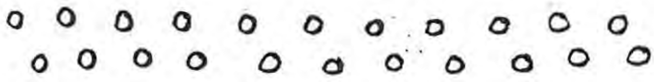
(3)


(5)


(8)


Incomplete figure

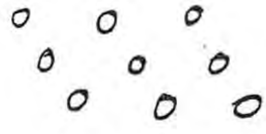
(1)  (2)  (6 or fewer dots or columns of circles)

(2)  (2 instead of 3 rows of circles)

(4) 

(5) 

Do Not Score

(2)  (statement was made that *S* did not want to draw the other circles)

5. RETROGRESSION ***Definition.*** Substitution of a more primitive gestalt form than the stimulus.

Figures. Score for all figures except 4 and 6.

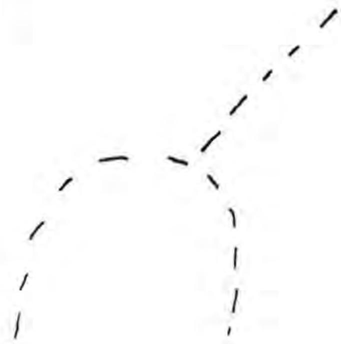
Loops for circles (if persistent)

(2)

Dashes for dots (if extreme and persistent)

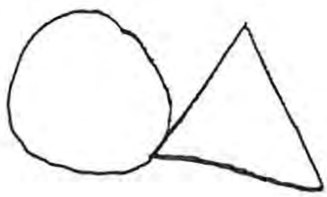
(1)

(5)



Triangle for diamond or hexagon

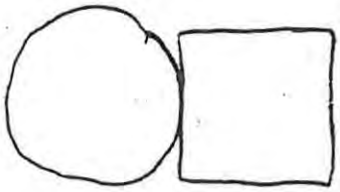
(A)



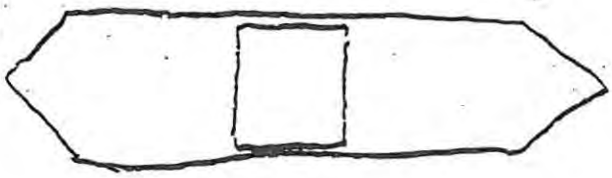
(7)

Square for diamond

(A)



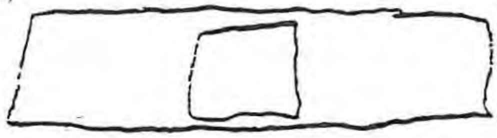
(8)



Rectangle for hexagon

(7)

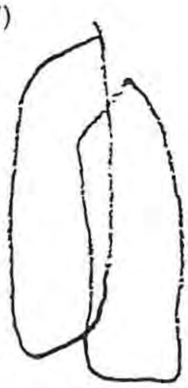
(8)



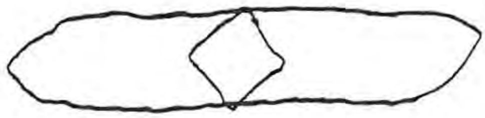
Do Not Score

Do not score if curves are substituted for angles or angulation of bottom of hexagon on figure 7 is omitted.

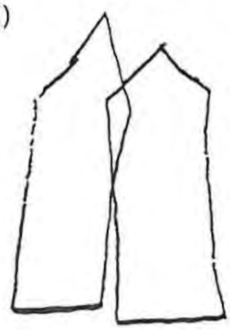
(7)



(8)



(7)

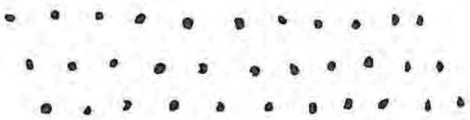


6. PERSEVERATION There are two kinds of Perseveration errors. If both occur, this error is still only scored once.

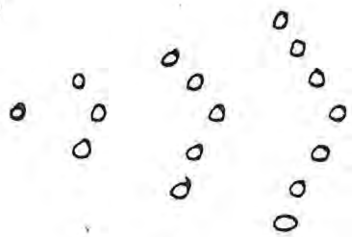
Type A *Definition.* Inappropriate substitution of the features of a preceding stimulus, such as replacing the circles of figure 2 with the dots of figure 1 (must have made dots, not circles on figure 1); replacing the dots of figures 3 and 5 with the circles of figure 2 (must have made circles on figure 2 and dots on 1).

Figures. Score only for figures 2, 3, and 5.

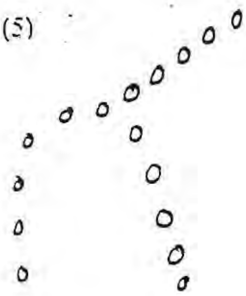
(2)



(3)



(5)

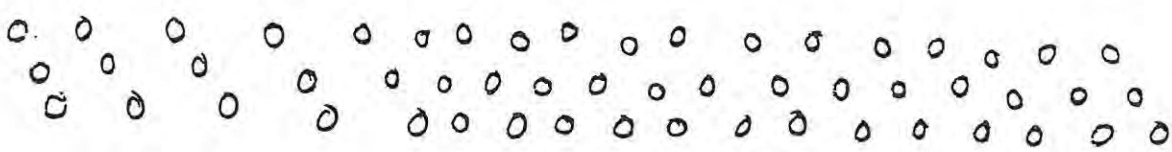


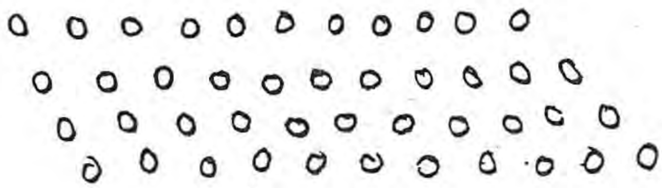
Type B *Definition.* Intradesign perseveration or continuing to draw a figure beyond the limits called for by the stimulus. For figure 1, 14 or more dots must be present; for figure 2, 13 or more columns of circles.

Figures. Score only for figures 1, 2, and 3.

(1)

(2)



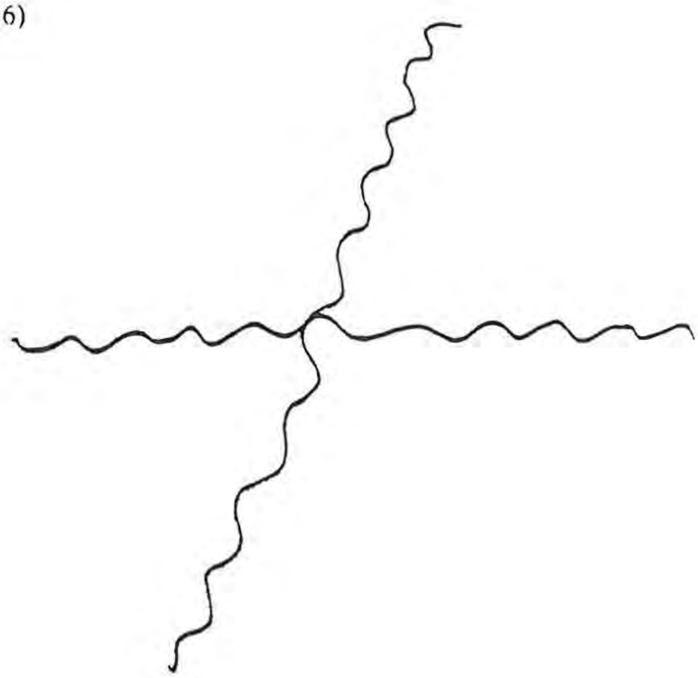
(2)  (an added row of circles)

(3)  (an added row of dots)

Do Not Score

(1)  (circles for dots on figure 1, score Simplification)

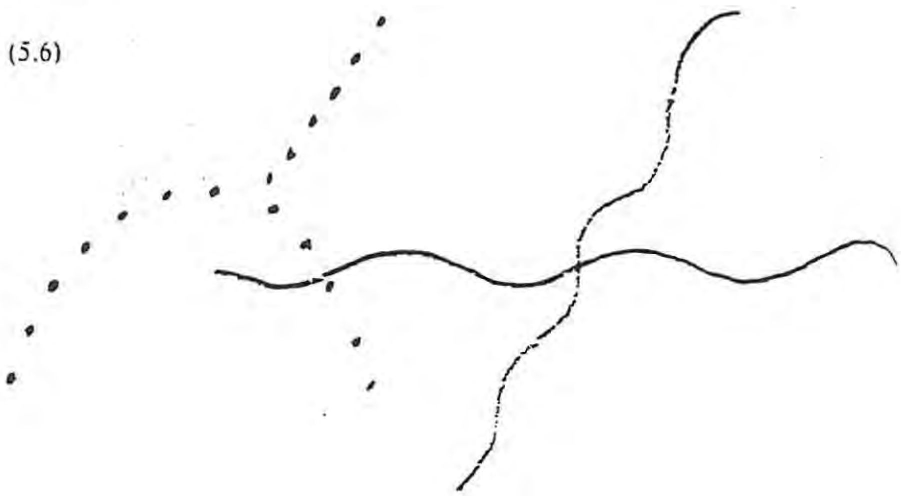
(6)



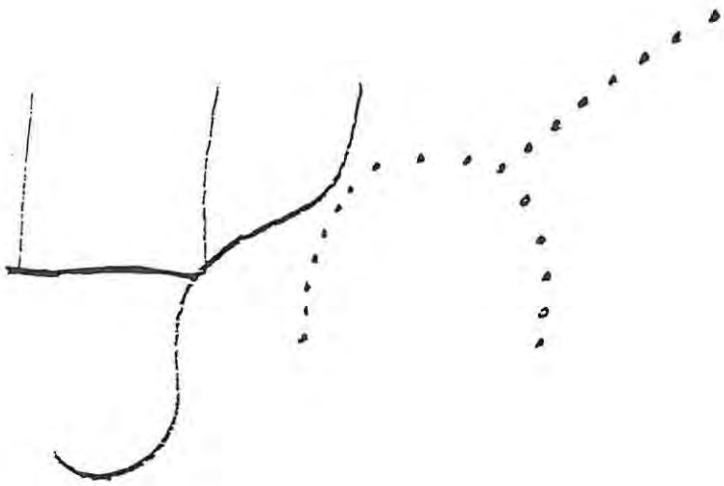
7. COLLISION OR COLLISION TENDENCY *Definition.* One figure is drawn as touching or overlapping another figure (collision) or is drawn within ¼ inch or less of another figure but does not touch (collision tendency).

Figures. Score for all figures.

(5.6)

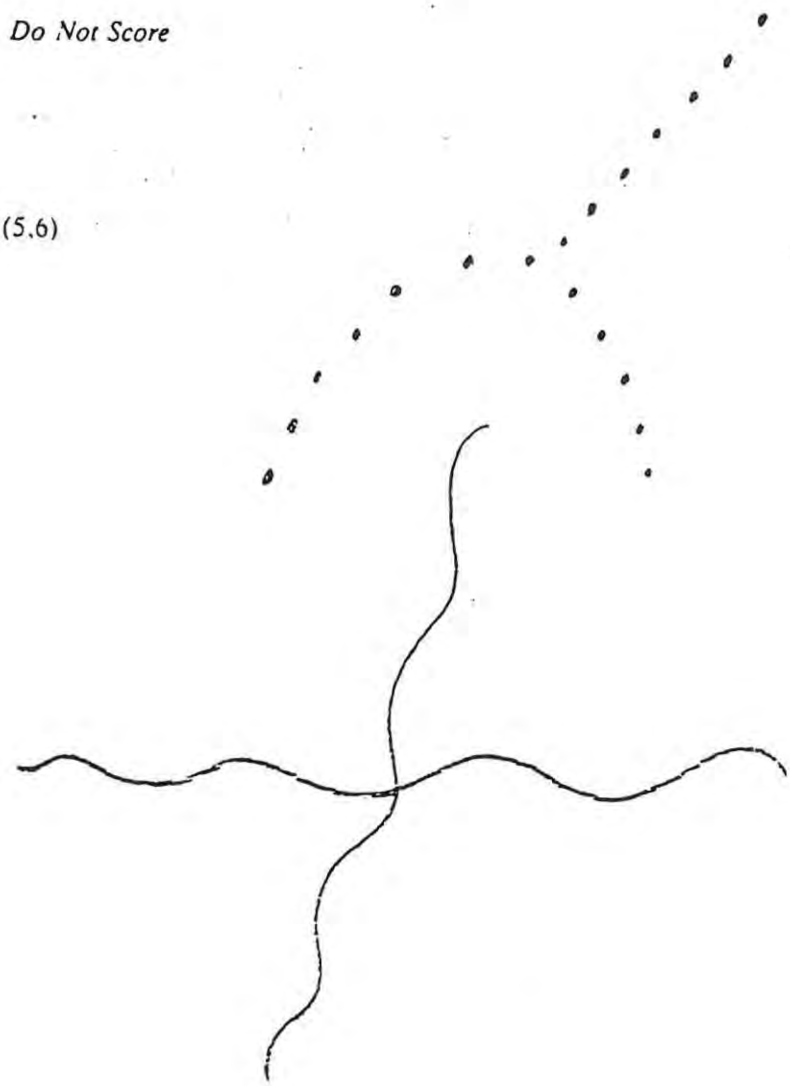


(4.5)



Do Not Score

(5.6)



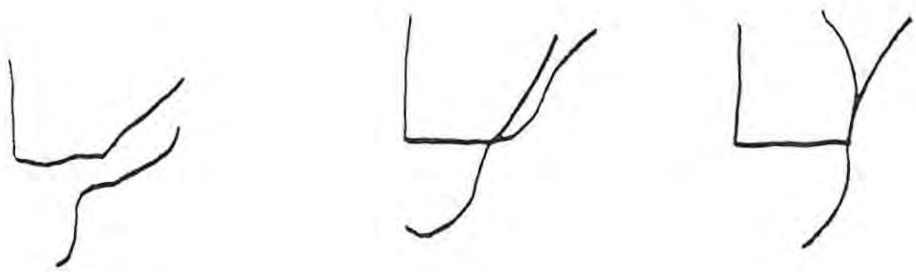
8. IMPOTENCE ***Definition.*** Behavioral or verbal expressions of inability to draw a figure correctly (often accompanied by statements such as "I know this drawing is not right but I just can't make it right").

Figures. Score for all figures.

Repetitious drawings or numerous erasures of figures with similar inaccuracies

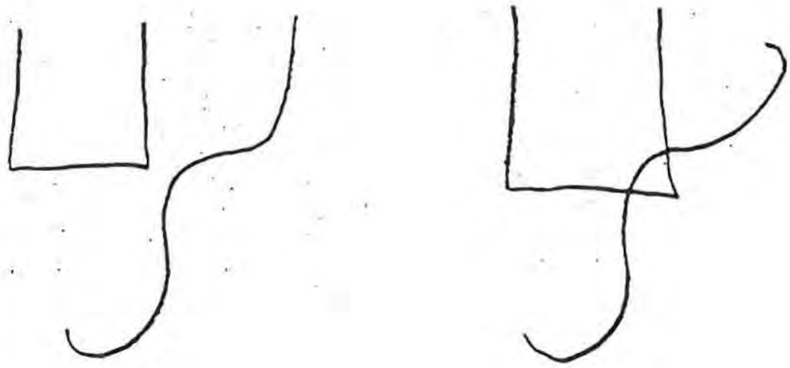
(A)

(4)



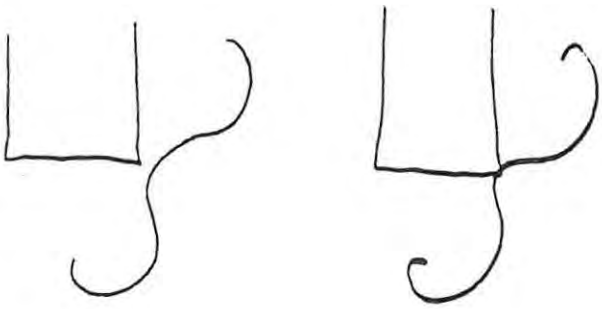
S realizes that an error has been made and tries to correct it unsuccessfully or expresses inability to correct it

(4)



Do Not Score

(4)



(second attempt corrects the error)

9. CLOSURE DIFFICULTY

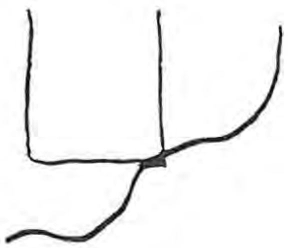
Definition. Difficulty in getting the joining parts of figures together or getting adjacent parts of a figure to touch. If figures are more than ⅛ inch apart at joining point, score Simplification.

Figures. Score only for figures A, 4, 7, and 8.

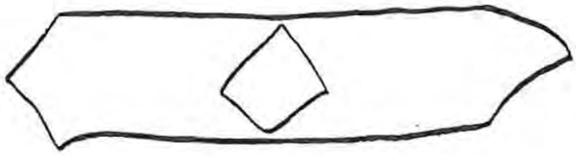
Consistent but not significant joining problems on 2 out of these 3 figures

(A)

(4)

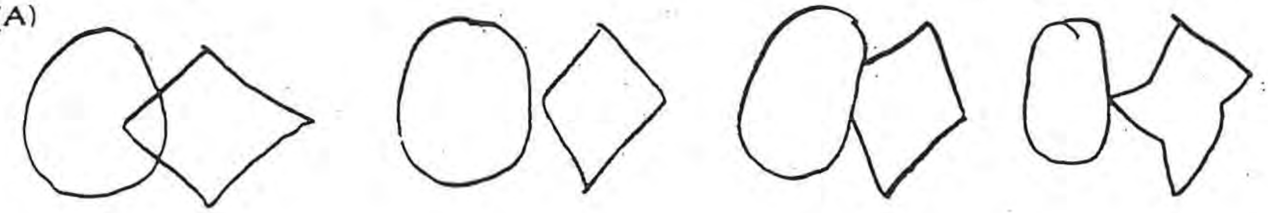


(8)

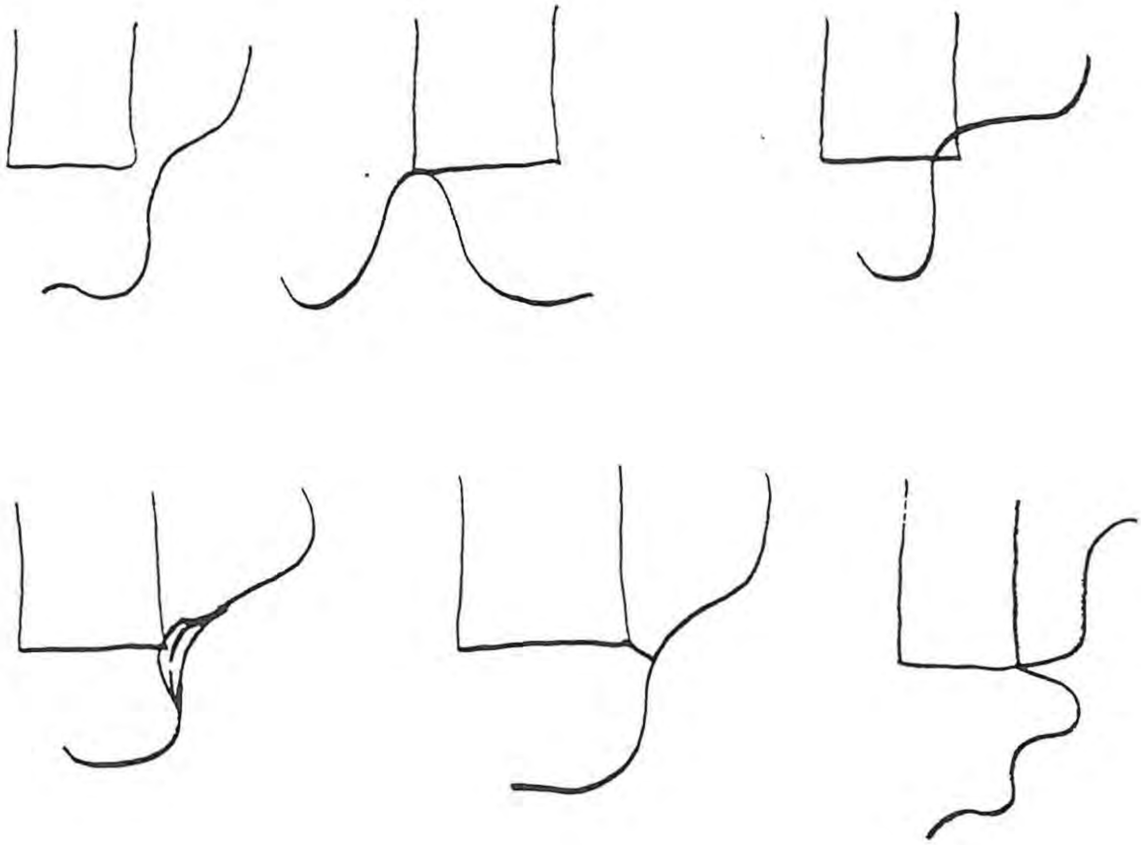


One *significant* problem with closing circles or figures or joining adjacent parts of a figure

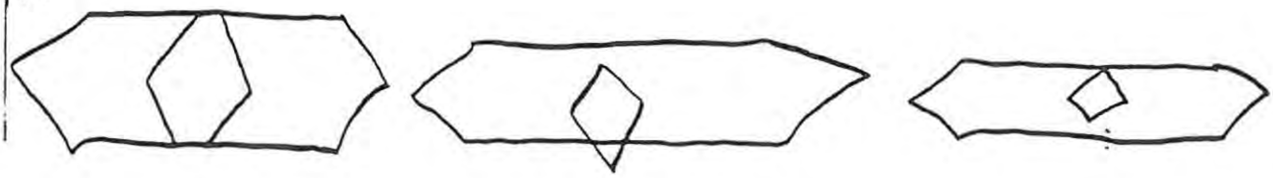
(A)



(4)



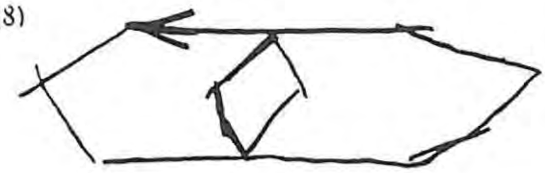
(8)



Marked and persistent gaps, overlap, redrawing, sketching, erasures, increas pressure at points where parts of the design join one another

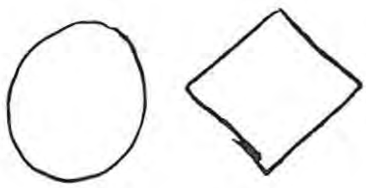
(4)

(8)



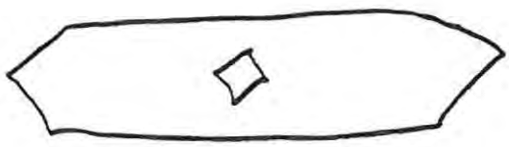
Do Not Score

(A)



(parts of figures are more than 1/8 inch apart, score Simplification

(8)



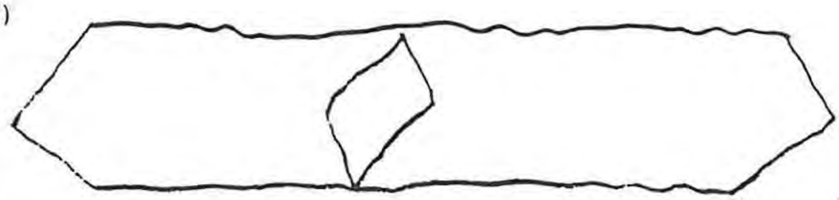
10. MOTOR INCOORDINATION

Definition. Irregular (tremor-like) lines, especially with heavy pressure. Behavioral observations are important for scoring this error. Be sure *S* is drawing on a smooth surface.

Figures. Score for all figures.

(1)

(8)



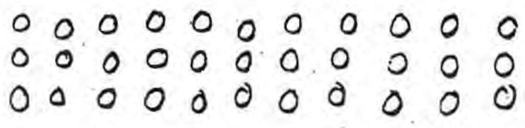
11. ANGULATION DIFFICULTY

Definition. Severe difficulty in reproducing the angulation of figures.

Figures. Score only for figures 2 and 3, but especially figure 2.

Failure to reproduce angulation of a figure

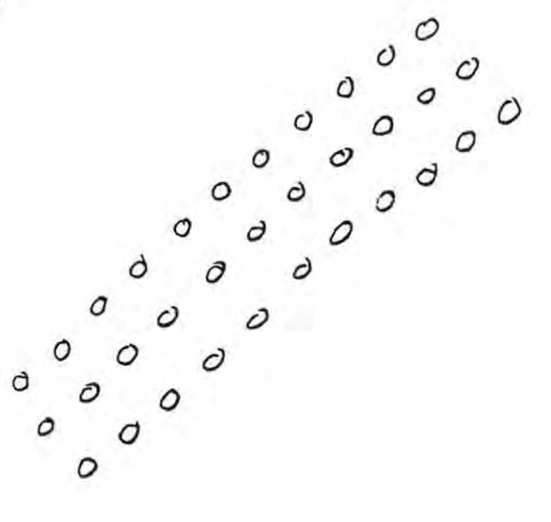
(2)



(3)

Angulation of the whole figure 45° to 80° rather than parts of a figure (but not by greater than 80°, which would be Rotation)

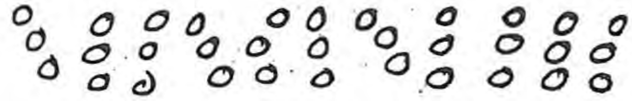
(2)



(3)

Variability of the angulation of more than half the rows of circles of figure 2

(2)

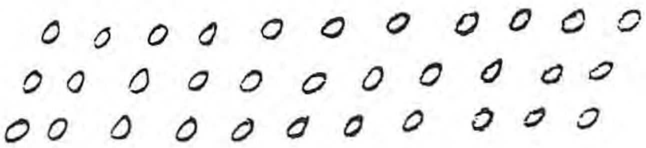


Do Not Score

(3)

(figure 3 should be scored leniently because its angulation is especially hard to reproduce)

(2)



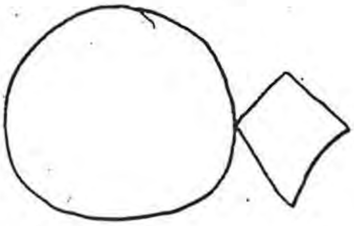
(reversal of angulation on figure 2, score Rotation)

12. COHESION ***Definition.*** Isolated decrease or increase in size of figures. Score very conservatively. This error is most frequently overscored.

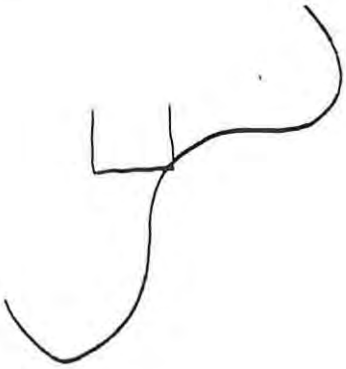
Figures. Score for all figures.

Decrease in the size of *part* of a figure by more than ⅓ of the dimensions used in the rest of the figure

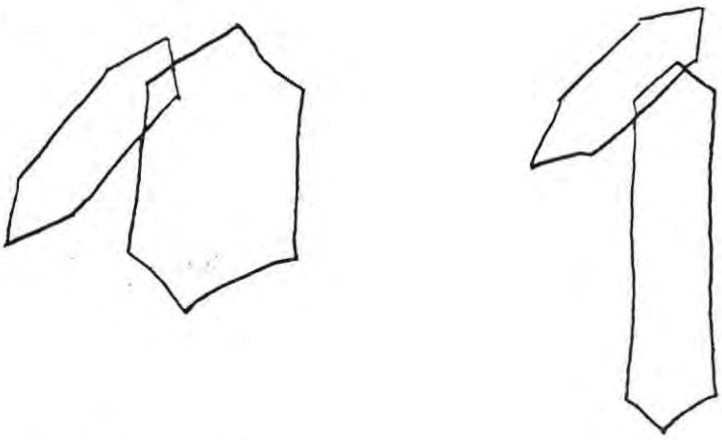
(A)



(4)



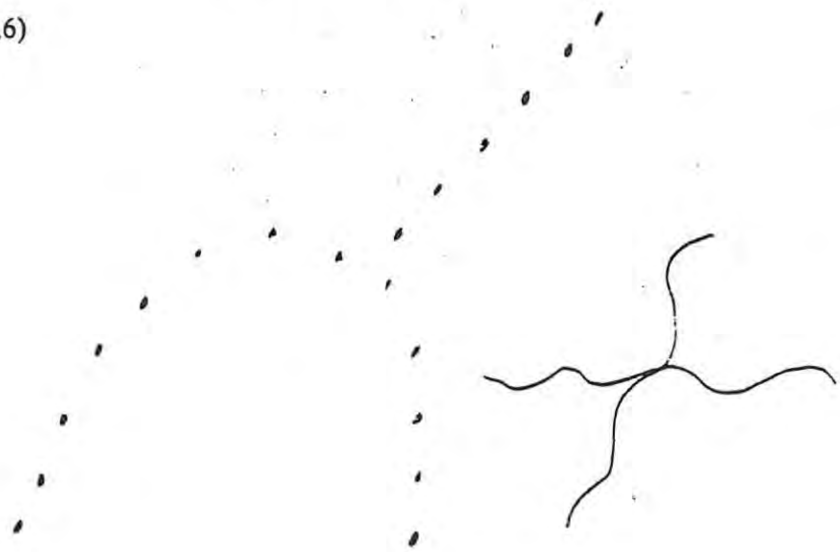
(7)



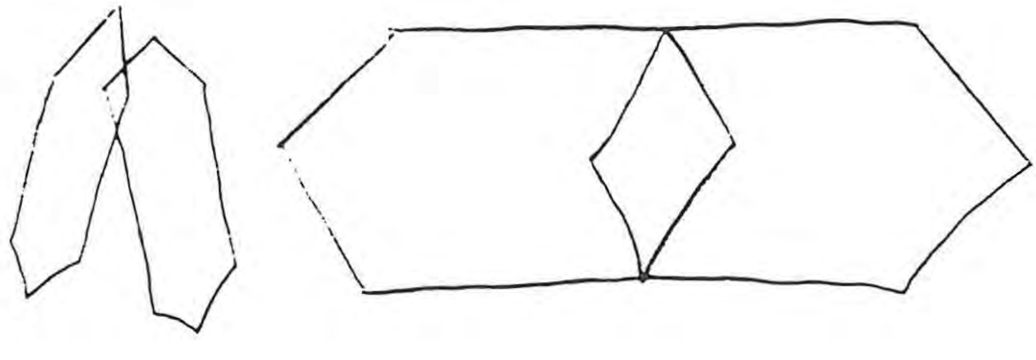
(8)

Increase or decrease in the size of a figure by ⅓ of the dimensions used in the *other drawings* (*not* compared to the size of the stimulus cards). Exclude parts of drawing that are larger due to Perseveration.

(5.6)



(7.8)



APPENDIX 3: LACKS'S (1984) OBSERVATIONAL CRITERIA (EXPANDED FOR THE GROUP TESTING OF LOW-EDUCATED ADULTS)

Lacks's observational criteria (1984, p.24) assist in alerting the tester what significant factors to look out for during observation of the test. The tester must endeavour to establish whether performance is a true reflection of the subjects' ability; i.e., whether errors are due to perceptual-motor impairment or to other extraneous factors. Lacks (1994) stresses the need of the tester to concentrate on the subjects during testing. Lacks (1984, p.24) observation criteria are as follows:

- Evidence of fatigue
- Insufficient attention to stimulus
- Extremely rapid and careless execution
- Extreme care and deliberation
- Dissatisfaction expressed over poorly executed drawings (client's reported efforts to correct these are not successful)
- Poor motor coordination or hand tremor
- Rotation (indicate which figure)
- Other
- Time[1]

Two other observational criteria were added by this author as already discussed (see 5.1.2: p.93 and 5.1.3: p.94).

- Anxiety
- Eyesight - apparent difficulty seeing the figures (e.g. may needs glasses)

---

1. This is an optional scoring error which is not used in this research, as already discussed (see 5.1.4, p.94).

APPENDIX 4: TESTING ADMINISTRATION

Testing instructions for the BGT-*c*

1. The nine Bender (1946) designs are pasted on 4 x 6 inch cards.
2. Subjects are tested in a group of 8 without a research assistant, or 14 with a research assistant.
3. Each subject is given a 2B pencil with an eraser.
4. The tester must ensure that the writing surface will is smooth (in order that an irregular writing surface does not simulate the effect of motor incoordination).
5. Plain white A4 paper is used and subjects receive a small pile each so they may use as much paper as they wish.
6. No ruler or other mechanical aids may be used.
7. Screens between each desk are used or alternatively, a reasonable space between the desks is ensured.
8. Behaviour is observed according to Lacks's (1984) observational criteria (see Appendix 3: p.257).
9. Approximate times for the completion of each test are noted.
10. The time lapse from completion of the BGRT-*i* to the start of the BGRT-*d* is 10 minutes.
11. The turning of the cards or paper is discouraged. Where this is done, the tester reorientates the card or paper. If a subject is insistent, the baseline on the page is marked.

Testing instructions for the BGRT-*i*

The cards are collected and the subjects are asked: "Will you please now draw as many of the designs you can remember from those you have just copied".

Testing instructions for the Goodenough Draw-A-Man Test

Subjects are asked, "Please draw the best man that you can".

Testing instructions for the BGRT-*d*

Subjects are asked, "Will you please draw as many of the designs which you have already drawn, again. This is the last part of this test."

APPENDIX 5: SUBJECT SHEET

Subject sheet for the group administration of the expanded BGT

Time started________________						
Subjects	Finish BGT-*c*	Finish BGRT-*i*	Start BGRT-*d* at:	Start BGRT-*d*	Finish BGRT-*d*	Observations
50*	8.16	8.28	8.38	✓	8.43	Impotence Fig. 3.

* Data for subject No. 50 have been filled in as an example

APPENDIX 6: DETAILS AND EXPLANATIONS OF PROCEDURES FOLLOWED

A: Using numbers to facilitate the group testing of low-educated adults

This method was specifically developed by this author for the group administration of low-educated adults as a number of subjects could not write their own name on each protocol or forgot to do so. Writing a number on a protocol each time one is removed from a subject is rapid and avoids the protocols from getting muddled. In addition, also to avoid confusion the following codes were used; 'c' for the BGT-*c*, 'i' for the BGRT-*i* and 'd' for the BGRT-*d*. This ensured that even if protocols did get muddled, they were easily sorted out.

Before testing began, a subject sheet (see Appendix 5: p.259) was drawn up with each number representing a subject. The time testing began was noted at the top of the page, and then the time at which each subject completed each test was simply written in. As soon as the BGT-*i* was completed, the tester added 10 minutes to indicate the time at which the BGRT-*d* should begin, and again marked the sheet once the subject was actually instructed to commence. The actual time period was worked out later, as there is insufficient time during testing to compute this. When working with a research assistant in a bigger group, one person organized the completed protocols, controlled the timing and made observational notes, while the other issued administration instructions to the subjects.

B. Masking the embarrassment subjects felt about inadequate writing skills

Some subjects experienced considerable embarrassment about their low educational level or lack of schooling. For this reason, some union leaders stipulated that no educational limits be put on volunteers. For example, during testing some subjects were embarrassed about not being able to write at all or sufficiently to record the information required. Thus, during the testing, efforts were made to mask those who could not write by asking anyone with questions or problems to put up their hands. The researcher(s) then discreetly wrote down the information required.

In many cases, participants volunteered, but then verbalized highly anxious feeling about not being able 'to write properly'. This anxiety was counteracted by reassurances and boosting the work and the abilities of the subject. For example, a group of construction workers was particularly anxious until the researcher pointed out that if she tried to build a wall, she would be fired rapidly. The subjects found this highly amusing and were then quite happy to go ahead with the testing procedure.

APPENDIX 7: CHECKLIST

Scoring tables for Lacks's (1984) system

CHECKLIST

1. Rotation
2. Overlapping difficulty
3. Simplification
4. Retrogression
6. Perseveration
7. Collision or tendency
8. Impotence
9. Closure difficulty
10. Motor incoordination
11. Angulation difficulty
12. Cohesion
13. Time ________
14. Behavioural observ.

SCORE __________

CHECKLIST

1. Rotation
2. Overlapping difficulty
3. Simplification
4. Retrogression
6. Perseveration
7. Collision or tendency
8. Impotence
9. Closure difficulty
10. Motor incoordination
11. Angulation difficulty
12. Cohesion
13. Time ________
14. Behavioural observ.

SCORE __________

CHECKLIST

1. Rotation
2. Overlapping difficulty
3. Simplification
4. Retrogression
6. Perseveration
7. Collision or tendency
8. Impotence
9. Closure difficulty
10. Motor incoordination
11. Angulation difficulty
12. Cohesion
13. Time ________
14. Behavioural observ.

SCORE __________

CHECKLIST

1. Rotation
2. Overlapping difficulty
3. Simplification
4. Retrogression
6. Perseveration
7. Collision or tendency
8. Impotence
9. Closure difficulty
10. Motor incoordination
11. Angulation difficulty
12. Cohesion
13. Time ________
14. Behavioural observ.

SCORE __________

CHECKLIST

1. Rotation
2. Overlapping difficulty
3. Simplification
4. Retrogression
6. Perseveration
7. Collision or tendency
8. Impotence
9. Closure difficulty
10. Motor incoordination
11. Angulation difficulty
12. Cohesion
13. Time ________
14. Behavioural observ.

SCORE __________

CHECKLIST

1. Rotation
2. Overlapping difficulty
3. Simplification
4. Retrogression
6. Perseveration
7. Collision or tendency
8. Impotence
9. Closure difficulty
10. Motor incoordination
11. Angulation difficulty
12. Cohesion
13. Time ________
14. Behavioural observ.

SCORE __________

APPENDIX 8: MEAN SCORES ON THE LACKS (1984) SCORING SYSTEM BY AGE AND EDUCATION

Study	Age	Educational Level														
		2,5 years			5,5 years			8,5 years			> 9 years			8,8 years[1]		
		n	*M*	*SD*	*n*	*M*	*SD*	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Dyall (1993)	7,6-10,1	37	-4,30	1,51												
Dyall (1993)	10,9-12,9				38	-3,00	1,29									
Dyall (1993)	13,3-16							39	-1,66	1,19						
Lacks (1984)	17-24										2	-	-			
Lacks (1984)	25-34										3	-2,00	1,00			
Lacks (1984)	35-44										10	-2,70	2,16			
Lacks (1984)	45-54										37	-3,08	1,38			
Lacks (1984)	55-59										23	-3,61	1,62			
Lacks *et al.* (1982)	60-64													22	-3,23[2]	1,67
Lacks *et al.* (1982)	65-69													85	-3,00[3]	1,97
TOTAL		37	-4,30	1,51	38	-3,00	1,29	39	-1,66	1,19	75	-3,20	1,51	107	-3,11	1,82

1. Ranges from no formal education to completed college;
2. 23% had scores of 5 or more;
3. 21% had scores of 5 or more;
- Indicates less than 3 subjects.

APPENDIX 9: CORRESPONDENCE

9A: Letter to John Wiley and Sons Inc.

9B: Letter of permission from John Wiley and Sons Inc.

Poste Restante
Grahamstown,
6140
South Africa.
Tel: (0461) 25141
Fax: (0461) 24738

5 January, 1995

Wiley & Sons
New York
Fax: (212) 850 6008

Dear Mr Kelly,

Following our telephone conversation yesterday, I hereby request permission to use an excerpt from a book published by your company.

The book is "Bender Gestalt Screening for Brain Dysfunction" by Patricia Lacks (1984). It is a Wiley-Interscience Publication. The reason for my request is that I am currently completing a Masters' Degree in psychology on the Bender Gestalt Test. I have applied the Lacks' scoring system to children and very low- and uneducated adults. This appears to be the first such application as no published literature can be found on this scoring system used for these groups and increases the applicability of the test. I have developed a scoring system for the Bender Gestalt Recall Test, based on Lacks's and another author's scoring systems. Since the understanding of the Lacks system is fundamental to my research, it would be appropriate to include this in my thesis. The pages I wish to use are 83-110. The manuscript is an unpublished one, and at the very most about 20 copies would be made, but it is more likely be about eight copies.

As a former journalist I am fully aware of the copyright legislations. Although I plan to publish my work at some stage, I would request permission again. The current granting of permission applies only to this unpublished thesis. Obviously your company would receive full acknowledgement.

Secondly, I would greatly appreciate it if you were able to give me Ms Lacks's address, as I would like to correspond with her.

Finally, as discussed, once I have completed my thesis, I will send a copy to your Vice-President: Publications, Mr Stephen Kippur, with a view to his perusal of the material for publication.

Many thanks for your consideration.

Sincerely,

Kate Dyall (Mrs)