

**THE EFFECTS OF A MODIFIED TAT IN  
SOUTH AFRICA: A COMPARATIVE  
ANALYSIS**

Ruth Scott

A minor dissertation submitted in partial fulfilment of the  
requirements for the award of the degree of  
MASTER OF ARTS (CLINICAL PSYCHOLOGY)  
in the Faculty of Humanities  
UNIVERSITY OF KWAZULU-NATAL  
2008

Supervisor: Prof. Douglas Wassenaar

This work has not been previously submitted in whole, or in part, for the award of any degree. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, and has been cited and referenced.

Signature: \_\_\_\_\_ Date: \_\_\_\_\_

## Abstract

This study proposed to assess the effects of a modified TAT in a multi racial South African sample. More specifically, it aimed to measure how different race groups responded to a variation of the TAT, as opposed to the original Murray TAT. It was hypothesized that a racially inclusive TAT, with more contemporary backgrounds, would enhance the narratives of South African respondents, and more especially those of black respondents. A new set of five TAT pictures (e-TAT) was designed, based upon five of the original Murray TAT (M-TAT) cards. As much as possible, the ambiguous style of the cards was left intact, but the characters (all originally white) were changed to represent African/black, Indian and coloured people as well. A factorial design was used to compare responses on the Murray TAT to responses on the experimental TAT in a sample of 207 first year psychology students at the University of KwaZulu-Natal. Each student was given a mean word count score. A randomly selected sub-sample of 40 students received a quality rating derived from Gerver's *scoring level of response* (as cited in Coleman, 1947), intended to measure the quality of their protocols. The statistical analyses revealed that (1) There is no statistically significant differences in the length of protocols between the M-TAT and the e-TAT. (2) There was no statistically significant difference in the length of stories between the race groups. (3) On the smaller sub-sample of participants who received quality ratings, black participants scored significantly higher on the e-TAT than on the M-TAT. (4) White participants scores on the M-TAT were significantly higher than black participants scores on the M-TAT. While the results are not yet conclusive, they are encouraging, and it is suggested that future research in this area is needed.

## **Acknowledgements**

Thank you to my supervisor, Prof. Douglas Wassenaar, whose support, insight, and patience throughout the duration of this project were invaluable. Thank you also to Warren King, Frank Bokhorst, Michael Quail and Molelekoa Kometsi for their time and assistance. Credit is due to Lyn de La Motte for illustrating the modified TAT cards.

## Table of Contents

<b>ABSTRACT</b> .....	<b>I</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>II</b>
<b>TABLE OF CONTENTS</b> .....	<b>III</b>
<b>CHAPTER 1 INTRODUCTION AND LITERATURE REVIEW</b> .....	<b>1</b>
1.1 Overview .....	1
1.2 A Note on Terminology .....	3
1.3 Literature Review.....	3
1.4 Historical overview .....	4
1.4.1 A brief history of psychological assessment.....	4
1.4.2 Difficulties associated with test standardization in South Africa .....	6
1.4.3 A retrospective overview of projective personality testing .....	8
1.5 The Thematic apperception Test.....	12
1.5.1 The Construction of the tat and underpinning rationale .....	12
1.5.2 Factors determining identification .....	14
1.6 Validity, Reliability and Norms .....	15
1.6.1 An overview of validity and reliability.....	15
1.6.2 Validity .....	17
1.6.3 Reliability.....	18
1.6.4 The dangers of Eurocentric norms.....	20
1.7 A Review of Previous Modifications.....	22
1.8 Rationale for a contemporary South African TAT .....	24
<b>CHAPTER 2 METHODOLOGY</b> .....	<b>27</b>
2.1 Sample.....	27
2.2 Instruments.....	27
2.2.1 Construction of the Experimental TAT .....	27
2.3 Experimental Design.....	29
2.3.1 Research Hypotheses .....	30
2.4 Procedure .....	31

2.4.1 Ethical Issues .....	31
2.4.2 Data collection .....	31
2.4.3 Scoring Procedures .....	32
<b>CHAPTER 3 RESULTS.....</b>	<b>35</b>
3.1 Overview .....	35
3.2 The effect of TAT and Race Group on Word Count .....	36
3.3 The effect of TAT and race group on mean quality score .....	37
3.4 In Summary.....	41
<b>CHAPTER 4 DISCUSSION.....</b>	<b>43</b>
4.1 Theoretical Overview.....	43
4.2 Implications of Results: .....	45
4.2.1 Results for word count.....	45
4.2.2 Results for productivity .....	46
4.3 'Quantity is not quality' .....	49
4.4 The use of aggregates.....	50
4.5 Limitations and Suggestions for Future Research .....	51
4.6 In Conclusion .....	53
<b>REFERENCES.....</b>	<b>54</b>
References: Primary Sources .....	54
References: Related publications .....	54
<b>APPENDIX A.....</b>	<b>59</b>

## List of Tables

Table 1. Sample Demographics .....	35
Table 2. Descriptive Statistics for Mean Number of Words Produced.....	36
Table 3. Analysis of Variance for Mean Number of Words Produced.....	37
Table 4. Sample Demographics .....	38
Table 5. Descriptive Statistics for Mean Quality Score.....	38
Table 6. Analysis of Variance for Mean Quality Score.....	39

## List of Figures

Figure 1. Means plot of the interaction between Race and version of TAT on mean productivity .....	40
Figure 2. Means plot of the interaction between Version of TAT and Race on mean productivity score.....	41
Figure 3. Murray TAT Card 1.....	59
Figure 4. Experimental TAT Card 1 .....	59
Figure 5. Murray TAT Card 2.....	60
Figure 6. Experimental TAT Card 1 .....	60
Figure 7. Murray TAT Card 4.....	61
Figure 8. Experimental TAT Card 3 .....	61
Figure 9. Murray TAT Card 6GF .....	62
Figure 10. Experimental TAT Card 4 .....	62
Figure 11. Murray TAT Card 9GF .....	63
Figure 12. Experimental TAT Card 5 .....	63

# **Chapter 1**

## **Introduction and Literature Review**

### **1.1 Overview**

Psychologists in psychodiagnostic, forensic, industrial, educational and other contexts are frequently required to assess an individual's personality, usually with the purpose of understanding the uniqueness of their character (de Bruin, 2001). The Thematic Apperception Test (TAT), a projective technique designed to elucidate dimensions of personality via open ended narratives, is one of the most popular assessment measures next to the Minnesota Multiphasic Personality Inventory and the Rorschach (Rossini & Moretti, 1997). Projective tests are based loosely on the psychoanalytic assumption of projection, in that respondents are thought to project unconscious aspects of their personality onto the test question, interpretable by an analyst (Colman, 2001).

Developed by Morgan and Murray of the Harvard Psychological Clinic in 1935, the TAT has been (Bailey & Green, 1977), and continues to be, one of the most widely used projective techniques in clinical practice (Narron, 2005). The technique, however, is said to be as controversial as it is popular, with particular concern directed at its questionable validity across ethnic groups (Hibbard, Tang & Latko, 2000). Despite perennial predictions of its demise, the TAT is said to elicit a person's "interpersonal attitudes, subtle psychopathological processes and motivation for overt behavior with remarkable sensitivity" and is used extensively irrespective of a client's referral, history or cultural background (Rossini & Moretti, 1997, p.393).

This research project investigates dimensions of the cross-cultural applicability of the original Murray TAT (1935) for both black and white urban South Africans. Measuring differences between the quality of narrative productions between the old TAT and an experimental version, it aims to explore the effect of different versions of the TAT on narrative length and quality. It also investigates the interactions between race and culture relative to the type of TAT pictures presented in the South African context. Essentially,

this research explores whether there is both need and justification for the development of a new South African or racially and culturally inclusive South African TAT.

The structure of the thesis is outlined below.

#### *Literature Review*

This section will include a brief history of psychological and projective testing followed by an outline of the relevant literature including previous attempts to modify the TAT and the results of these modifications. Political motivation for changing the tests, specifically in South Africa, will be discussed with a view to justifying this particular method of modification and understanding of culture.

#### *Research methodology*

The sample, procedure (data collection and measurement) and statistical analyses used in the research will be discussed and presented in this section.

#### *Research results*

ANOVA tables will be presented and expanded upon. All relevant results and assumptions necessary for the statistical procedures will be tabulated and briefly explained.

#### *Analysis of results*

An in depth analysis of the results as related to the original hypotheses and previous research will be critically discussed. The results will also be looked at from a cultural psychology perspective and the implications of the research will be explored.

#### *Summary and conclusion*

In this section, a summary of the main findings will be presented. Implications for further research will be discussed, essentially for the purpose of either advocating or opposing the construction of a new, South African TAT.

## **1.2 A Note on Terminology**

Students were asked to identify themselves according to the population group categories defined by Statistics South Africa. These are African/black, coloured, Asian/Indian or white. For the sake of brevity, throughout the thesis, the term 'black' will be used inclusively to refer to those who identified themselves with the population categories African/black, coloured or Asian/Indian

To avoid confusion, throughout the thesis 'African/black' is referred to as 'African'.

While the name of the TAT suggests it is a test (Thematic Apperception Test), preference is given to the term 'technique' as 'test' mainly refers to structured, objective and standardized measures that are used for gathering specific information (Foxcroft & Roodt, 2001). Because of the broader connotation, the TAT will be referred to as a 'technique' throughout the thesis.

## **1.3 Literature Review**

Psychological tests, including projective techniques, are frequently used to aid psychologists in making decisions and recommendations that can have significant implications in the diagnosis and treatment of patients in various settings. Within these contexts, psychologists are often required to assess an individual's personality, identifying their strengths, weaknesses and typical ways of relating to the world (de Bruin, 2001). Other decisions where psychological tests are often applied include: educational placement, custody and employment (Williams, 2006).

This study hopes to be of interest to clinicians using projective tests as well as to projective testing research. Mirroring Narron's (2005) assertion, the lack of material on the outdated (and, arguably, culturally biased) nature of the TAT is considered a "disservice to the psychological community as a whole" (p.13). While much attention has been focused on the validity of the TAT, few researchers have addressed the need to

update this highly used technique<sup>1</sup>. As a prelude to the focus on projective testing, with particular reference to the TAT, a broad outline of the history of psychological testing is discussed. A detailed account is beyond the scope and relevance of this research, yet it is illustrative to situate projective testing within the wider discipline of psychological assessment. This discussion will draw heavily on Foxcroft and Roodt (2001).

## **1.4 Historical overview**

### **1.4.1 A brief history of psychological assessment**

A general concern with understanding and predicting the behaviour of 'others' is age old and probably stems back to ancient prehistoric times. In addition to intuitive and unsystematic methods that have most likely always been part of interpersonal relations and human development, other organized procedures attempting to make sense of 'selves' and 'others' such as astrology, palmistry and humorology have existed since pre-Biblical times (Goodstein, 1971). While some of these methods are still practiced today, they are regarded as pseudo-scientific approaches due to the perceived incongruence with modern science and criticism directed at the lack of empirical demonstration (Goodstein, 1971). The development of psychological testing as a method to answer similar questions and understand interpersonal behaviour is fundamental to the emergence of psychology as a profession. Perhaps the greatest contribution to the development of psychology as a science, was the development of objective psychological testing. Underlying the formative scientific methods in psychology was measurement, behavioral assessment and empiricism (Foxcroft & Roodt, 2001).

The first milestone in the history of psychological assessment can be traced to Huarte's book published in 1698, *The Trial of Wits*, where it was proposed that people's differences could somehow be measured. It was suggested that personal individual attributes and capacity should be assessed and matched to appropriate occupations or educational settings. Behavioral rating scales, dependent on individual behavioral

---

<sup>1</sup> As explained in the note on terminology, the TAT is referred to as a technique throughout the thesis.

observations, were among the first uniquely psychological measurement tools to emerge (Foxcroft & Roodt, 2001).

By the late 19th century, behavioral and personality assessment was respected and treated with similar reverence to 'the experiment', that had traditionally been the domain of physical science. Rigorous control and standardized conditions became hallmarks of psychological assessment, and psychology began to prosper as a scientific profession (Tomkins, 1947). The term psychometrics, coined by Wolff, was used throughout the 18th and 19th centuries. However, only towards the 20th century did the significance of psychometrics shift towards measuring individual cognitive ability and assessing personality processes (Foxcroft & Roodt, 2001).

The rapid progress of psychological assessment in the 20th century has been attributed to advances in statistical methods for analysis of data as well as an increase in the theories of human behaviour that served as the foundation underlying the construction of new measures (Foxcroft & Roodt, 2001). In 1905, Binet and Simon developed the first measure of intelligence that was intended to identify mentally handicapped individuals requiring special education. Ten years later, with large numbers of military recruits needing to be assessed during World War I, the need for individual psychological assessment was even more urgent. This in turn broadened the scope of testing, and tests of achievement, aptitude and personality started to flourish (Foxcroft & Roodt, 2001). During this boom period of testing, especially by the 1920's, several psychologists in the United States were associating themselves with the applied aspects of the field and concentrated on gaining recognition, both within the discipline, and from the general public (Goodstein, 1971). Knight Dunlap (1971) writes of this climate and how it provided plenty of opportunity for frauds, whose greed, enhanced by the gullibility of the public, cast the profession of psychology and psychological testing into temporary disrepute.

By World War II however, the use of psychological assessment was firmly re-established. In 1943, the development of the Minnesota Multiphasic Personality Inventory (MMPI), introduced a new era in personality assessment. The MMPI was the

first objective and structured personality test that emphasized the use of empirical data in determining test results (Foxcroft & Roodt, 2001). In applied settings, psychological testing became one of the most important functions of psychologists. Furthermore, in 1954, the American Psychological Association (APA) declared that the right to use these measures was restricted to Clinical Psychologists (Foxcroft & Roodt, 2001).

In response to rapid globalization, coupled with more critical thinking in psychology, by the close of the 20th century, multiculturalism became a focus of thought in testing and assessment. Owing to the fact that by the beginning of the 21st century most available measures had been developed in the US and UK, attention turned to cross cultural adaptation (Foxcroft & Roodt, 2001). Test adaptation refers to the process of making a test more suited, relevant and applicable to people within a certain context while retaining its original meaning and aim. The intention is to create fair assessment where bias is removed and validity increased (Kanjee, 2001). South Africa has a rich and controversial history of test adaptation with particular reference to test standardization<sup>2</sup>. Unfortunately however, test adaptation and standardization, in a political climate of discrimination based on apartheid ideology, increased bias and cast doubt on the validity of tests for all South Africans. In the next section, issues around standardization in South Africa are briefly reviewed.

#### **1.4.2 Difficulties associated with test standardization in South Africa**

The trajectory of psychological testing in South Africa is similar to that seen in Europe and the U.S. However, a fundamental difference to be noted is the political and social context within which this development took place (Foxcroft & Roodt, 2001). The history of psychology and psychological assessment in South Africa is intimately intertwined with the history of South Africa's racial segregation and the unequal distribution of resources characteristic of the apartheid era (Hook, 2004).

---

<sup>2</sup> In psychometrics, test standardization refers to establishing norms (the average range of scores) for the population the test intends to measure (Colman, 2001).

The standardization and construction of measures in South Africa (even preceding 1948) was driven by political ideologies that attempted to identify innate distinctions between races, where one race group was deemed superior over another. This is especially seen in the history of intellectual testing where tests adapted from the U.S. and Europe were standardized for white populations only (Foxcroft & Roodt, 2001). Because the tests were normed on exclusively white populations, and owing to inferior schooling, teaching methods and a widespread unfamiliarity with western testing techniques among African, Indian and coloured populations, research inevitably showed that mean scores on intelligence tests for black<sup>3</sup> school children were inferior to the mean scores of white school children. Theorists used this dubious research to 'prove' the innate inadequacy of black, South Africans. Differing cultural environments, access to resources and the effect of malnutrition on intelligence, all contributed to the gross test bias encountered, but were for the most part ignored. Biased research results continued to be used to reach conclusions that perpetuated racism and separate development in South Africa (Foxcroft & Roodt, 2001).

Current theorists acknowledge that applying tests of cognitive ability from one ethnic group to another, without appropriate standardization is highly problematic and potentially discriminatory (Shuttleworth, Kemp, Rust, Muirhead, Hartman & Rudolph, 2004). Rather than differences in test scores being understood as a difference in ethnic attributes per se, the phenomenon should be seen as the effects of learned socio-cultural factors that happen to characterize particular groups of people within similar contexts. Complicating this issue further is that 'ethnic' or 'cultural' groups may be more different within their group than they are similar in terms of socio-cultural characteristics. According to Shuttleworth et al. (2004, p. 904) "socio-cultural influences encompass a number of closely inter-related variables that are difficult to separate, including language usage and reading ability, level and quality of educational attainment, socio-economic status, home and schooling socialization experiences". Rapid globalization has caused widespread movement within population groups and communities which further confounds the notion that ethnic or racial groups are bounded and static entities. Multiple

---

<sup>3</sup> As explained in the note on terminology, the term black is used inclusively to refer to African, Indian and coloured people in South Africa.

factors therefore are creating immense challenges for the standardization of testing in South Africa.

The following section discusses the history of projective testing in slightly more depth. Again, a full account is beyond the scope of this study but some background to the context within which the TAT developed is of interest.

### **1.4.3 A retrospective overview of projective personality testing**

In 1879 the first free association experiment was performed by Francis Galton (Tomkins, 1947). He was both experimenter and subject – discovering that recurring associations over a four month period could be mostly traced to his youth and formative experiences, while those that occurred only once were related to more recent occurrences. Galton acknowledged the diagnostic potential of his findings but pursued these no further (Tomkins, 1947). Experimental psychology continued to focus attention on understanding human ‘thought patterns’ and by the turn of the century, the education system called for the more ‘complex’ task of measuring general intelligence (Tomkins, 1947). The potential of Galton’s findings was to remain untapped for many years.

According to Tomkins (1947) the motivation to study personality mostly came from mental hospitals in Europe and the U.S.. The mental testing movement however was dominated by educators and psychologists who narrowly concerned themselves with intellectual function, leaving personality to fall into the hands of psychiatrists. Psychology only fully embarked on the theory of personality from 1907 on publication of Brittain’s “A study of Imagination” (as cited in Tomkins, 1947). Brittain presented a collection of 9 pictures to young children and asked them to write stories about the scene presented. These narratives were then analyzed according to their imaginative quality, length and their use of moral, social and religious elements. Sex differences were found where girls’ stories revealed interest in the preparation of food and clothes while boys were more interested in the consumption of food. Furthermore, the girls’ stories had more social, moral and religious elements than the boys’ accounts. The following year a similar

study by Libby (as cited in Tomkins, 1947) investigated the relationship between feelings and imagination with a developmental focus. Through the use of pictures and narrative, Libby found that the latent meaning in the imagination of younger children was the same as in adolescents.

The concept of projection was first developed by Sigmund Freud, initially with regard to paranoia (in what we now term Delusional disorder). While Freud himself never provided a detailed analysis of the concept, in brief, projection is now understood as a defense mechanism in which a person's unbearable emotions, impulses, or thoughts are falsely attributed to other people (Colman, 2001).

The original technique presented by Brittain remained unexploited for 25 years until it was rediscovered by the psychiatrist Schwartz who developed the Social Situation Picture Test (Tomkins, 1947). The test was developed for use with 'juvenile delinquents' because time pressures in the court clinic did not allow for the rapport needed in a full psychiatric interview. The test consisted of a series of 8 pictures depicting scenes familiarly encountered in the histories of young offenders. Young male subjects were asked to describe what the boy in the picture was thinking, and then later asked what he would do and think if he was the boy in the card. The test was never widely used by psychologists however, owing to the specificity of the scenes to young criminals and anticipated difficulties with standardization (Tomkins, 1947).

Original use of the terms 'projective testing' and the 'projective hypothesis' is attributed to US psychologist, Lawrence K. Frank (1890-1968) who, in 1939, published an article in the *Journal of Psychology* entitled 'Projective Methods for the Study of Personality' (Colman, 2001). While, according to Goldberg (1971), there has never been a singularly satisfactory theory or framework explaining the operation of projective techniques, the projective hypothesis is commonly used as the theoretical rationale underlying the use of projective tests. Presented simply, the projective hypothesis states that when an individual imposes meaning or order on an ambiguous stimulus, his response is a 'projection' of his inner most emotions, drives, beliefs, thoughts, and desire (Goldberg, 1971).

Frank (1971) understood an individual's personality to be the complex result of a dynamic interaction between culture and the psyche. Personality for Frank was a process of mutual influence whereby an individual organises and understands his environment through the lens of an inner private world. That private world is however dependent on and created through cultural agents such as family, environments and other external events. He therefore believed that in order to access the personality through testing, one needed to evoke the process of personality as it had been and continued to develop. By forcing the individual to integrate and understand an ambiguous yet cultural stimulus representative of the external world, the inner private world is called upon (whether this is consciously chosen or not). These techniques lay in sharp contrast to the traditionally popular quantitative measures (Frank, 1971).

During, and particularly after, World War II, projective techniques became increasingly popular in the emerging profession of clinical psychology. Some of the techniques had already been in use for some time. Rorschach's inkblots, for example, had been available since the early 1920s; word association was championed by Jung much earlier; and the TAT was first introduced into the literature in 1935. However it was only after the war that 'projective techniques' really gained momentum (Rabin, 2001).

According to Rabin (2001) the term 'projective testing' gave a number of existing clinical diagnostic methods, previously excluded from traditional psychometric batteries, a theoretical home of their own. Also included were some common psychometric tests such as the Stanford-Binet and the Wechsler-Bellevue tests. All were constructed with the view to yield broader personality investigations than had previously been possible through the traditional strictly quantitative measures such as Mental Age, IQ, Scatter, and so forth.

In 1947, *Case Histories in Clinical and Abnormal Psychology* by Burton and Harris was published. The text was intended to present the state of the art case histories culminating at the end of the war. In an introductory chapter to the book, Henry A. Murray (1947) stressed the limitations of the field, asserting that it lacked depth and sophistication in its understanding of individuals. He associated these limitations with the narrow conceptions

of clinical psychology common of the time. This sentiment was echoed at the 1949 Boulder conference which delineated national U.S. standards for clinical psychology training in universities (Rabin, 2001).

In response to these criticisms and the perception that psychological understandings, training and literature are limited, attempts were made to broaden the scope of the discipline in a number of different ways. These steps allowed for the tentative integration of projective techniques into the orthodox and mainstream agendas of clinical psychology. First, there was the cautious and gradual movement of specialized workshops and seminars (particularly on the Rorschach) from the more elite clinical settings, into the academy and graduate programmes. Many universities began to accept, with varying degrees of enthusiasm, the usefulness of some of the projective techniques, despite their lack of statistical validation and support. Second, with the growing number of clinical training programs in several universities, several texts and guides for the use of various specific techniques were introduced. Finally, the introduction of a standard curriculum for university training in clinical psychology, as suggested in the report of the Boulder Conference, further encouraged the making explicit of the various diagnostic methods that showed off the 'new look' in the field (Rabin, 2001). A review by Bell (1948) and an edited volume by Abt and Bellak (1950) were seminal precursors.

Projective tests are now defined as any of a variety of personality assessments that require free responses or associations to ambiguous stimuli such as ink blots, incomplete sentences or pictures (Colman, 2001). These tests require the assumption of projection, the premise being, that patients will reveal intolerable and therefore repressed aspects of their personality by 'projecting' these unconscious processes through their responses. Theorists hold, that through projection, people disown or disavow certain aspects of themselves by assigning those aspects to the external environment. People or even objects can be at the receiving end of a projection, where they become a receptacle for unwanted or intolerable aspects of the projector's personality (Corey, 2005).

Projective assessment techniques are often interpreted or 'scored' in a more intuitive impressionistic way than structured personality inventories and questionnaires. The

potential advantage of such technique is that the chances of socially desirable response sets are said to be reduced and more holistic results achieved (de Bruin, 2001). Disadvantages however are seen in the plethora of literature on their dubious validity and reliability (to be discussed in section 3.1).

Perhaps one of the most remarkable features of projective testing today, is that in spite of extensive reports condemning the techniques as theoretically unsound, unreliable and invalid, they continue to be the psychological instruments used most often (Hock, 1999). The two most well known and popular projective tests are the Rorschach inkblot and the Thematic Apperception Test and both have been pivotal in the history of psychology (Louw & Edwards, 1997). The following section will concentrate solely on the TAT. It is argued that the historical background of the technique contributes to its perceived cultural bias and also that the underpinning rationale reveals some profoundly Eurocentric assumptions.

## **1.5 The Thematic apperception Test**

### **1.5.1 The Construction of the tat and underpinning rationale**

The original TAT (M-TAT) consists of a set of 31 cards, depicting black and white illustrations or photographs. The pictures were collected from old magazines and media of the day by Murray and Morgan in the U.S. during the early nineteen thirties (Tomkins, 1947). The pictures are a mixture of some relatively clear everyday scenes and other more vague indistinct representations. All the pictures however are thought to be ambiguous and tell an incomplete story, facilitating unique or personal interpretation on behalf of the test taker. The M-TAT pictures have a particularly European quality (Bullard, 2005) which is small wonder owing to the time and place they were collected and created in. Moreover the characters are all white and most often dressed in European clothing characteristic of the period.

Test takers are asked to look at the picture and tell a story about the given scene so that it contains a beginning, a middle and an end. Where possible, test takers are also

encouraged to furnish the characters in their stories with the emotions they think appropriate to the subject. Murray (1943, p. 3) suggests instructing adolescents and adults of average to high intelligence as follows:

This is a test of imagination, one form of intelligence. I am going to show you some pictures, one at a time; and your task will be to make up as dramatic a story as you can for each. Tell what has led up to the event shown in the picture, describe what is happening at the moment, what the characters are feeling and thinking; and then give the outcome. Speak your thoughts as they come to your mind. Do you understand? Since you have fifty minutes for 10 pictures, you can devote about 5 minutes to each story. Here is the first picture.

Stories are supposed to elicit basic needs, motives, perceptions of the environment and feelings that are only partially conscious (Louw & Edwards, 1997). Originally, subjects were asked to interpret the action of each picture and guess preceding events and subsequent outcome. Through experience however it was learned that more revealing narratives resulted when test takers were asked to tell a dramatic story (Tomkins, 1947). Morgan and Murray (cited in Hock, 1999, p. 270) wrote "The purpose of this procedure is to stimulate literary creativity and thereby evoke fantasies that reveal covert and unconscious complexes". In trying to explain ambiguous human behaviour in the 'other', it was believed that the test taker or patient would become less self conscious and less defensive, consequently revealing their own repressed wishes, fears and experiences (Hock, 1999). The theory underpinning the TAT, at its most basic level, is that human behavior is fuelled by unconscious forces. This view requires an implicit acceptance of psychoanalytic psychology and it is held that unconscious conflicts must be exposed and interpreted for the accurate diagnosis and treatment of psychological problems.

Because the pictures were chosen and developed with the aim of stimulating the respondent's own conflicts and dynamics, Morgan and Murray felt it was important that there should be at least one person in every picture that the subject could easily

recognize. Subsequently, in one of their earliest studies of the TAT they discovered that the stories their sample came up with were inspired by four main sources including: popular fiction; lived experiences involving friends or relatives; events in the participant's own life and lastly the test taker's unconscious and conscious fantasies. Their second important discovery was that participants were most likely to project their own psychological dynamics into their stories when they themselves identified with characters in the cards (Hock, 1999). The TAT has been used in the exploration of a wide variety of personality areas including child development, social attitudes, assessment of military personnel and culture (Tomkins, 1947).

### **1.5.2 Factors determining identification**

If identification with the cards or characters within them is so central to the TAT's utility, then a discussion regarding the requirements for such identification is necessary. According to Rotter (1946) the respondent is unlikely to identify with a character in every card; however, certain conditions increase the likelihood of a participant's identification (thus projection) with the scene. Firstly, Rotter (1946) asserts, when other conditions are held constant, the respondent is most likely to identify with a character of the same sex and similar (or previous) age. For example a young female adult would be likely to identify with a young female adult or girl in the card but be unlikely to relate to an older female or a male of any age.

Secondly, Rotter (1946) goes on to suggest that a respondent is unlikely to identify with a character who engages, or is seen to be engaging in, socially deviant behavior in the opinion of the test taker. This view could be extended to say that the respondent is also unlikely to identify with a character engaging in culturally deviant or foreign behavior from the test taker's perspective. Rotter (1946) gives an example saying that it is improbable for the respondent to identify with a character who is seen to be cruel, unjust, perverted, stupid or mean in the participant's eyes. Similarly, translating this in the South African cultural context one might assert that a test taker from an isiXhosa or isiZulu cultural background may not identify with a character who is acting in culturally

incongruent ways. Certain behaviours perceived normal from a Western cultural standpoint may be considered rude, disrespectful or inappropriate from a different cultural position. Taking a well known example such as eye contact – while making eye contact might imply honesty and directness for a Westernized individual, for some people with an IsiXhosa or IsiZulu cultural upbringing, this kind of eye contact with an elder is deemed rude and disrespectful (Naidu, 2008; Makhale-Mahlangu, 1996). It is important however to be aware that exceptions do occur, for example if the test taker needs to construct him or herself as an ‘outsider’ or rebel of some sort.

Narron (2005) argues that the outdated nature of the cards elicits narratives that are reflective of the historical period rather than ‘true’ projection. Using the common distinction between ‘experience far’ and ‘experience near’ stimuli he asserted that because the cards are so old – they constitute ‘experience far’ stimuli which are not ideal conditions for projection to take place. Modern and hence experience near stimuli are thought to be needed for meaningful engagement and identification with the cards.

The condition most pertinent to the present work that Rotter (1946) mentions is with regards to history. He says that a participant is most likely to identify with a character whose history or background is most similar to his or her own. While in Rotter’s argument it appears that he is mostly discussing history and background in terms of occupation, education and family, it is the theoretical standpoint of this thesis that cultural identification is equally, if not more, important. It is argued that culture informs these background or historical characteristics and serves as a lens through which people construct, understand and present their experience of self.

## **1.6 Validity, Reliability and Norms**

### **1.6.1 An overview of validity and reliability**

“Generations of psychologists, including personality and clinical psychologists, have been trained with a deeply ingrained assumption that projective techniques are inherently

invalid and unreliable” (Westen, Lohr, Silk, Gold & Kerber, 1990, p. 362, as cited in Cramer, 1999, p. 90).

Since 1938, and particularly in the last 5 decades, controversy has surrounded the reliability and validity of the TAT (Lilienfeld, Wood & Garb, 2000; Tomkins, 1947). While a full review of the plethora of literature on this topic is beyond the scope and relevance of this work, a brief summary of current thinking is outlined. This section will draw heavily on Lilienfeld et al.’s (2000) article titled “The scientific status of Projective Techniques” as well as an article by Cramer (1999).

Lilienfeld et al. (2000) reiterate a commonly cited paradox, namely, that on the one hand there has been a litany of assessment researchers publicly denouncing the validity and reliability of most projective techniques, while on the other, widespread use of these techniques has continued unabated. According to a report by Durand, Blanchard, and Mindell (1988, cited in Lilienfeld et al., 2000) 65% of directors responsible for clinical psychology internships believe in the importance of formal training in projective techniques. Moreover, the authors add that many registered clinicians suggest that these assessments are invaluable to their daily practice. A brief synopsis of concerns directed at the legitimacy of the TAT in terms of validity and reliability, as well as some counter arguments grounded in narrative theory follows.

Lilienfeld et al. (2000) begin their largely negative declaration on the scientific status of the TAT by outlining the many problems found with the research and literature on the psychometric properties of this projective instrument. These criticisms mostly concern the unregulated variability of administration and scoring of the TAT. While Murray (1943) advocated two sessions, where clinicians use 20 of the cards with themes corresponding to the patient’s hypothesized difficulties, this recommendation is almost never followed in contemporary practice. There is said to be substantial variation in both the number of cards used and the order or themes of the cards presented. According to the authors, this in turn means that there is little consistency regarding administration of the TAT in published research, making claims on validity and reliability dubious to say the least. Similarly, there is considerable inconsistency in clinicians and researchers use of

scoring systems, as a variety of quantitative scoring schemes have been developed.<sup>4</sup> Few clinicians regularly use the same schemes, and more confounding is the fact that most psychologists in their clinical work, use no scheme at all, and prefer to score the narratives intuitively (de Bruin, 2001; Lilienfeld et al., 2000). According to Lilienfeld et al. (2000) the vast range of these schemes, not to mention the variety of stimulus materials on which they are founded, all but precludes a methodical enquiry into the TAT's potential psychometric properties.

### **1.6.2 Validity**

Incremental validity is held in high regard by Lilienfeld et al. (2000), owing to the time and effort it takes to administer a TAT. If, for example, the same amount and quality of information can be obtained from self report measures or demographic data, then they assert that routine use of the TAT in clinical settings is hard to justify. According to these authors, research shows that the incremental validity of the TAT over and above information typically obtained from a full psychiatric interview has been unimpressive. Cramer, however (1999) argues that material obtained from a TAT protocol is by definition not available to consciousness. Self reports should thus not be correlated with, or used to validate information from a TAT. Nevertheless, a number of studies reviewed by Lilienfeld et al. (2000) lent evidence to their assertion, that adding TAT material to basic demographic data did not significantly improve the validity of a clinician's personality assessment.

Lilienfeld et al. (2000) are similarly concerned with treatment utility, and emphasize that if there is scant or no evidence suggesting that the projective instrument improves treatment results or therapeutic efficacy, then rationalizing their use is called into question. They go on to claim that no research supports treatment utility for any of the projective instruments. They call for more studies within the treatment context, challenging personality researchers to prove the projective test's contribution to treatment efficacy.

---

<sup>4</sup> Lilienfeld et al. (2000) cite Bellak (1975); Dana (1955) and Arnold (1962) as examples.

Lilienfeld et al. (2000) acknowledge some support for the construct validity of a few scoring schemes. Nevertheless, they assert, even these schemes lack adequate norms, evidence of test-retest reliability and validity across cultural groups. Furthermore, they questioned whether even the modest support for the construct validity of certain scoring schemes necessitates the TAT in itself, or if other methods could achieve similar and better validity. Hofer and Chasiotis (2004), however, were far more positive about construct validity, citing Spangler (1992) who found strong support for the validity of measures of nAchievement across 105 empirical studies. Hofer and Chasiotis (2004) thus concluded that "a carefully designed methodology in combination with an elaborated scoring system results in reliable and valid findings in Euro-American samples even though thematic apperception methods usually do not reach the levels of test criteria that objective methods do" (p. 225).

### **1.6.3 Reliability**

The TAT is repeatedly criticized for its poor reliability (inter-rater, test-retest and internal consistency) which is in turn said to constrain the test's validity. This point however has been disputed by theorists who assert that classical test theory is not suited to the TAT owing to an over-determined application of the reliability criteria (Hofer & Chasiotis, 2004). Lilienfeld et al. (2000) focused reliability critiques on three of the well known scoring systems of the TAT, which are (1) need based scoring schemes (McClelland, Atkinson, Clarke & Lowell, 1953, cited in Lilienfeld et al., 2000), (2) the assessment of Object Relations with the TAT (Westen, Lohr, Silk, Gold & Kerber, 1990, cited in Lilienfeld et al., 2000) and (3) the assessment of Defense Mechanisms with the TAT (Cramer, 1991, cited in Lilienfeld et al., 2000). According to Lilienfeld et al., the literature suggests that none of these scoring systems attains reliability satisfactory enough to justify their continued popularity and perceived authority.

Cramer (1999) argues compellingly that some measures of reliability are inappropriate for the TAT. In terms of internal consistency reliability, she argues that it is well known that different cards in the TAT were designed to elicit different themes and to assess

different areas of psychological functioning. Unlike items on a personality scale that are all intended to measure the same personality trait, there is no reason to anticipate that scores based on one picture will be identical, or even similar to those based on another.

Lilienfeld et al. (2000) take issue with the impact of this lack of inter-correlation between cards on the final result and interpretation of a TAT assessment. Supporting Entswile's (1972) view, referring to need based scoring schemes, Lilienfeld et al. (2000) contended that there is no use in obtaining aggregate scores when there is no internal consistency between the different cards of the test. In other words, because the cards are not correlated to each other and do not consistently measure the same attributes, an overall TAT score is essentially meaningless. Hibbard (2003) however replied, arguing that throughout the discipline of clinical psychology useful aggregates (often in the form of diagnoses) are obtained from different scales or symptoms. "Consider that each subunit of an aggregated group of predictors of a construct could be unrelated to the other, but when found in combination, they might well predict important variance in a construct" (Hibbard, 2003, p. 264). He used diagnostic systems as an example, saying that individual symptoms that are unrelated to each other become aggregated to form an overall picture or diagnosis. Similarly, when understanding personality using the Big Five personality factors, which are necessarily uncorrelated, a fully congruent picture can emerge from disparate parts.

Test-retest reliability has enjoyed similar controversy in the literature and typically, results are not promising (Lilienfeld et al., 2000). Cramer (1999) argues that there is difficulty inherent in measuring test retest reliability of TAT scores. This is for a number of reasons including, a reduction in novelty value on second presentation of the card, as well as the tendency on behalf of the test taker to feel compelled to tell a different story on second measurement (Winter & Stuart, 1977, as cited in Cramer, 1999). Studies have demonstrated that when participants are informed that they may tell identical or similar stories on subsequent assessment, nAchievement can increase from between .20 and .40 to approximately .60 (Lundy, 1985 as cited in Hofer & Chasiotis, 2004). Furthermore, test-retest reliability is difficult to measure meaningfully when taking cognizance that many features of personality (defenses for example) can change over time through

psychological development. While test retest reliability assumes that these measures are ultimately stable and constant, significant time lapses, especially in the event of therapeutic intervention could alter results significantly (Cramer, 1999).

Cramer (1999) suggests that inter-rater reliability might be the most appropriate measure of reliability for the TAT owing to the fact that it is an observational method rather than a psychometric test. Reliability in observational methods, Cramer (1999) argues, is based on agreement between two or more independent interpreters. In a review of the literature Hofer and Chasiotis (2004) report satisfactory levels of inter-rater reliability ( $r = .85$  or higher), particularly when elaborate coding systems are used.

In conclusion to her article, Cramer (1999) asserts that assessing the validity of the TAT is meaningless. Respecting that the TAT is a narrative, in Cramer's opinion, means acknowledging that these narratives are neither true nor false. Narrative theory interrogates meaning rather than fact and similarly recognizes that there are multiple meanings and multiple truths inherent in the stories of human's lives. The TAT "represents a construction of reality, not a reconstruction" (Cramer, 1999, p. 74).

#### **1.6.4 The dangers of Eurocentric norms**

Phenomenology in psychology concentrates on subjective descriptions of conscious experience, rather than specific or observable behaviour (Colman, 2001). While projective techniques, being largely phenomenologically based instruments, appear to offer holistic results that have tapped into universal personality phenomena, as Fanon and Geronimi (cited in Bullard, 2005, p. 235) stated, "the imaginary consciousness is certainly unreal, but it is fed by the concrete world". Therefore if the concrete and social world of the test taker is substantially different from the concrete and social world the test originated in, imaginary and emotional repertoires expressed are likely to be as substantially different, and hence misinterpreted.

Moreover, the scoring and interpretation of the TAT are reliant to an extent either implicitly or explicitly on the use of norms (Ehrenreich, 1990). If norms are understood as the way people 'usually' respond to or perform in a test, the necessary question is 'which people'? Whose 'normal' makes up norms and what are the implications for those who fall outside of those distributions? It is little surprise that the Murray TAT has a Euro-American normative standard. The standardization was done on college students from Harvard and because of who was encouraged to, or able to pursue tertiary education at Harvard in the 1930's, an affluent white male norm can be assumed (Bullard, 2005).

Traditional research and psychological practice that originated in the West are not neutral or value free but are constructed according to Western "sickness histories" (Gaines, 1992). How appropriate then is Western psychology for different cultural groups? If psychology is a cultural practice immersed in the ideal of rationality and individuals, how appropriate is it for cultural groups, who in their collectivist ideal, might value a spiritual or supernatural sense of the world (Bracken & Thomas, 2001)? What is to become of the mental health professions if there is no universal reality? Should there be different psychologies in different cultures? But what are cultures and are they synonymous with race or place or both? What about countries like South Africa that are said to be 'multicultural' and ethnically diverse?

These questions and criticisms however, are not new. As early as 1950 the TAT received criticism for being Eurocentric and therefore foreign to the non-Western subject it often attempts to assess (Bullard, 2005). Nevertheless, the Murray TAT has been used extensively across varying social populations with little critical reflection or caution in interpretation (Bailey & Green, 1977). Frantz Fanon, an avid critic of what he perceived to be the universalist agendas of Western psychiatry asserted that what is probably interpreted by the analyst as a restricted imagination or an inability to understand the ambiguous panoramic scenes, may be an incongruence between test material and the personality or life experience of the non-western test taker (Bullard, 2005).

## 1.7 A Review of Previous Modifications

Issues with standardization are certainly not unique to the TAT and in fact, attempts at cross cultural mental health care are almost as old as the discipline itself, both developing in the 19th century (Bullard, 2005). While the Murray TAT continued to be used extensively in the face of these critiques, and remains popular today, efforts to create culturally specific revisions of the TAT began soon after the original version. A wide variety of adaptations have been constructed for differing racial and age based groups (Monopoli & Alworth, 2000). These attempts however were not as noble as they claimed to be. According to Bullard (2005), colonial psychiatry used non-westerners' poor results on the original tests to 'scientifically claim' that the colonized were of inferior intellect and thus deserved culturally specific tests of a lower standard. South Africa in particular has had a rich history of TAT revisions most of which were used for testing 'non-white' candidates for employment (Bullard, 2005; Retief, 1987).

Much of the psychological testing was performed and designed under the auspices of the Institute for Personnel Research, founded and directed by Simon Biesheuvel. Pioneered by Lee, who in 1953 developed a TAT for use among Zulu speaking people (Ainsworth, 1959), these tests were widely employed in South African industry (Bullard, 2005).

"In short, Lee's TAT for African subjects recommends itself for...the investigation of culture personality interactions among African peoples who may be presumed to identify with near naked subjects" (Ainsworth, 1959, p. 313).

According to Retief (1987), Lee's TAT was developed from the fantasies of black men in a mental institution. Aside from the problems associated with norming a personality test on mentally disturbed inpatients, the rationales for the cards were not explained in the manual. Inferring what the cards are supposed to 'measure' is a highly problematic endeavor vulnerable to broader social dynamics that in South Africa at the time were likely to enforce cultural stereotypes and racism.

On publication of De Ridder's TAT series that appeared in 1961, a norm of cultural stereotyping inferring a 'primitive' and 'uncivilized' character type in black South

Africans continued. De Ridder's test was designed to select bus drivers during the apartheid regime in South Africa, but it was also used to make gross generalizations as to the character of South African blacks. For example, responses to a card depicting a hand holding money lead to his conclusion that all black Africans were obsessed with money (Retief, 1987).

The TAT-Z, designed for the evaluation of Zulu people, was published in 1976 by Erasmus from the Human Sciences Research Council. Again the purpose of the test was for selecting and placing non-whites in industry (Sikhosana, 1986). Once more the modified TAT relied heavily on cultural stereotypes, that not surprisingly, elicited stereotyped responses. Retief (1987) asserts that many of the cards were intended to assess black attitudes to white authority figures and would be expected to elicit mortification and anger in politically aware test takers. Any expression of anger towards the card however, was analyzed and the participant was penalized. In these cases, a much lower final score resulted. According to Bosman (1984, as cited in Sikhosana, 1986) by the mid 80's approximately 40 companies and state departments used the TAT-Z extensively and over 234 psychologists had been trained by the HSRC to administer and interpret the test. By October 1980, over thirty-three thousand TAT-Z booklets had been bought. "The fact that the cards are found to be useful should rather be seen as evidence of the pervasiveness of discrimination, and of the way in which discrimination is continued by interpreting tests in certain ways" (Retief, 1987, p.51).

A revision outside of South Africa worth discussing took place in the U.S. Thompson modified the Murray TAT for use with black Americans based on the assumption that identification with the characters on the cards facilitates more valid clinical personality evaluations (Ainsworth, 1959). His modification, unlike South African designs did not alter the background and stimulus setting of the scene and amounted to putting black skin and so called 'African American' features onto the original characters (Bailey & Green, 1977). Thompson's cards however were found to significantly increase the length of black American subjects' protocols as opposed to a control group who were given the original Murray set. His findings suggest that subject test congruence is important if one is to interpret TAT narratives with sensitivity. While this is an important finding, Retief

(1987) reminds us that constructs, signs and symbols are also powerful aspects of the stimulus material and should not be ignored, particularly in the South African context. The stimuli of the M-TAT reflect symbols and environments typically associated with Western culture in the early half of the 20th century which would be alien, or at best unfamiliar, to individuals with dissimilar cultural backgrounds (de Bruin, 2001).

In addition to the unsuitability of many of the original TAT cards for black South Africans, the stimuli may also seem outdated and unfamiliar for many white, coloured, and Indian South Africans. It seems that psychologists in South Africa are in need of a thematic apperception test that reflects the contemporary cultural complexity and richness of its society (de Bruin, 2001, p. 246).

### **1.8 Rationale for a contemporary South African TAT**

It could be argued that there is racism inherent most contemporary practice of the M-TAT in South Africa. Forcing the culturally foreign Eurocentric M-TAT on South Africans (especially black South Africans) may be both limiting, whereby participants do not resonate with and project onto the cards, and misleading, where an interpreter might understand shorter less emotion based narratives as defensiveness from the test-taker. Moreover, stereotyped revisions of the TAT for black populations (as outlined above) are equally racist and offensive for those living in culturally complex societies that involve a dynamic combination of traditional and modern influences. As Mkhize (2004) argues, one cannot be “narrowly focused on the role of local frameworks in the interpretation of human experience. Other frameworks cannot be ignored ...given that people do not live in impenetrable cultural enclaves” (p. 29). If one was to adopt a strictly relativist approach to psychological testing we would have to assume that ‘cultures’ are discrete and bounded entities that should be treated as distinctly separate. While with this approach, cultural context may be taken into account, it may also be reified, simplified and stereotyped – thereby homogenizing cultures and denying the fluid and dynamic nature of human essence. Swartz (1992) suggests that this kind of approach produces

beliefs and discourse reminiscent of the rhetoric used to justify apartheid in South Africa. Peoples' contexts include multiple cultural repertoires that are an assembly of local and westernized worldviews.

While this thesis employs quantitative method, the theoretical hypotheses have emerged out of post modern and social constructionist critiques of mainstream psychology. It challenges a pivotal assumption in traditional psychology that the route to the unconscious is not contingent on dynamic cultural meaning systems. Even universal phenomena are expressed culturally and it is hypothesized that these forms of expression or meaning frames are unlikely to be amenable to standardized or westernized scoring techniques. People make sense of themselves and their world through a complex cultural lens and so too are likely to express themselves according to that framework. The thesis is also premised on the goals of critical psychology. Critical Psychology aims to expose the oppression inherent in 'measuring' and categorizing all people according to Westernized norms and standards on the one hand and stereotyped views of culture on the other (Hook, 2004).

This project also hopes to show that while the focus is on projective testing, it's scope of meaning and relevance extends to the sociopolitical environment in South Africa. Psychology has long been intimately entwined with political agendas and historical forces (Foxcroft & Roodt, 1971). Inequalities present in wider social structures are often accurately mirrored in the practice of psychology especially seen in the South African apartheid climate. As Hook (2004, p. 136) puts it "the psychological repeats, reiterates and reinforces the political".

Owing to the perceived failure and racism inherent in previous attempts to revise the TAT for South Africans, this project attempts to begin a process of creating a more culturally valid set of cards integrating black and white people rather than just one or the other. It was hypothesized that urban South African students share cultural realities that transcend race, and that the Eurocentric Murray TAT panoramas are not suited and potentially restraining for black and white test takers alike.

This project could be classed as a theoretical indigenization whereby, it attempts to begin the process of developing a test consistent with the sociocultural realities of urban South Africans including “the use of locally derived reference systems, as well as borrowed theoretical frameworks that have been transformed to suit the needs of local populations” (Mkhize, 2004, p. 29).

In the following section on methodology, the sample, procedure (data collection and measurement) and statistical analyses used in this research will be discussed.

## **Chapter 2 Methodology**

### **2.1 Sample**

Participants were 207 first year psychology students at the University of KwaZulu-Natal. The age of the sample ranged from 17 to 28 years of age with a mean age of 19.21 (SD = 1.61). The first year course from which participants were drawn is divided into two classes who have the same lectures at different times each day, due to the large size of the class. One class was the experimental group (n=166) and the other the control (n=41). Students may choose which lecture to attend, so a random, self selected sample can be assumed, in that each student has an equal and independent probability of being in either condition. The condition (Murray TAT or Experimental TAT) was randomly assigned to the groups by the toss of a coin. The first lecture thus received the Experimental TAT (e-TAT) and the second, the Murray TAT (M-TAT).

### **2.2 Instruments**

The experimental materials consisted of cards 1, 2, 4, 6GF, and 9GF of the M-TAT. These cards form part of the standard set for both males and females (Bailey & Green, 1977), and in the author's opinion, were cards that especially highlighted the European quality of the technique (Bullard, 2005). Correspondingly these five cards were adapted to form the five e-TAT cards used in this study See (Appendix A).

#### **2.2.1 Construction of the Experimental TAT**

The experimental set of cards was developed by changing significant features on the five original M-TAT cards in an attempt to make them more culturally relevant and contemporary. The aim was that at least one significant feature would be changed in each

card, hypothetically lending a more culturally appropriate, modern, and 'experience near' (Narron, 2005) tone to the experimental version. Apart from the key features described below for each card, the background characteristics of the M-TAT were held constant in the e-TAT.

### **Card one (1)<sup>5</sup>**

In card one (1), the protagonist, a young boy, is made to resemble a coloured child by the slight shading of his skin and change of texture in his hair<sup>6</sup>. Furthermore, what was originally a violin has been changed to a guitar on the premise that guitars are more culturally resonant and common in South Africa than violins. This is not to say that most South Africans wouldn't recognise a violin, but that guitars are more common across a variety of populations and hence arguably, more familiar.

### **Card two (2)**

In card two (2), all characters were changed to resemble black African people, in skin colour, hair and facial features. In the one character (the woman leaning against a tree), clothing was changed to be similar to more modern variations of traditional African dress, including a head scarf. In addition, the horse in the picture was changed to a cow, which is argued to be more culturally familiar in South African contexts. Lastly, the barnyards in the background were modified to look like structures typical of a rural South African scene.

### **Card three (4)**

The modifications in card three (4), included changing the characters to resemble black African people in hair texture and style, skin colour, and facial features.

---

<sup>5</sup> Numbers in brackets correspond to the identifying numbers featured on the reverse side of the M-TAT cards.

<sup>6</sup> No features were intended to portray racial stereotypes. Care was taken to represent different South African population groups as accurately as possible, while maintaining subtlety as well as some of the ambiguity characteristic of the M-TAT cards.

#### **Card four (6GF)**

In card four (6GF) the characters remained white but their appearances were changed to bear more resemblance to modern South African white people. The hairstyle of the woman, which was particularly 'Victorian' in the M-TAT, was made into a style more commonly seen today. The older gentleman in the experimental card is now smoking a cigarette rather than a pipe, also in the hopes of appearing more modern, 'experience near' and familiar.

#### **Card five (9GF)**

In card five, both the characters were altered in an attempt to resemble young Indian girls. Their hair was made long and black and facial features were changed. Clothing was sketched to appear more modern but without the use of gross stereotypes or obvious, traditionally Indian, symbols.

### **2.3 Experimental Design**

The experimental design was a 2 (*version of TAT* – *e-TAT* versus *M-TAT*) X 2 (*race* – *White* versus *Black*) mixed factorial design with two levels in each independent variable. The independent variables were (1) *race* and (2) *version of TAT* administered. The two levels of the independent variable *version of TAT* were the *e-TAT* and the *M-TAT*. In turn the two levels of the independent variable *race* were (1) participants who identified themselves as *White*, and (2) participants who identified themselves as *Black*. The control and experimental groups were both mixed in race, and the version of TAT administered to the group was decided randomly by the toss of a coin.

The dependent variables were (1) *mean word count*, where each participant received an average score (number of words) across the five narratives produced (one narrative per card presented), and (2) *mean productivity rating*, where each participant is given a score between 1 and 5 according to the depth and quality of the narrative written, and the average across the five narratives is attained. Due to time constraints, the second

dependent variable, *productivity*, was only analyzed on a randomly selected sub-sample generated by STASTICA (2007)<sup>7</sup>. The sample consisted of 10 white participants from the control group (M-TAT), 10 black participants from the control group (M-TAT), 10 white participants from the experimental group (e-TAT), and 10 black participants from the experimental group (e-TAT).

### 2.3.1 Research Hypotheses

The research hypotheses can be summarised as follows:

1. The null hypothesis is that there will be no difference in story length for all participants between the e-TAT and the M-TAT.  
The experimental hypothesis is that there will be a significant difference in story length, with all participants producing longer stories on the e-TAT than on the M-TAT.
2. The null hypothesis is that there will be no difference in the quality of narrative for all participants between the e-TAT and the M-TAT.  
The experimental hypothesis is that there will be a significantly greater quality of narrative for all participants on the e-TAT than on the M-TAT.
3. The null hypothesis is that there will be no difference in story length for black participants between the e-TAT and the M-TAT.  
The experimental hypothesis is that there will be a significant difference in story length, with black participants producing longer stories on the e-TAT than on the M-TAT.
4. The null hypothesis is that there will be no difference in the quality of narrative for black participants between the e-TAT and the M-TAT.  
The experimental hypothesis is that there will be a significantly greater quality of narrative for black participants on the e-TAT than on the M-TAT.

---

<sup>7</sup> Statistica version 8 (2007) is the statistical software package written by StatSoft Power Solutions (1984-2008) that was used to do the data analysis for this thesis (<http://www.statsoft.com/>).

## **2.4 Procedure**

### **2.4.1 Ethical Issues**

The author of this thesis is bound by the UKZN Research Ethics Policy. Ethics approval was obtained only after a research proposal, addressing prominent ethical issues, was completed and accepted. The following ethical issues were addressed in the ethics approval proposal.

The autonomy of participants was respected and protected through the use of an informed consent form. The informed consent form specified, in plain language, the identity and affiliations of the researcher, the participants freedom to refuse or withdraw from the study, and the fact that a decision to participate in the study, or not to participate, had no bearing on their marks for the course. These issues were elaborated on and emphasized verbally before the participants began the task.

Anonymity was assured in that the participants were not required to provide their names or student numbers. Each participant had an identifying number for statistical purposes and the information will be destroyed after the thesis has been examined. The researcher treated the responses in a confidential manner. At no stage were participants' names used, nor was their personal information recognizable at any time.

It was explained to the participants verbally, and in the informed consent document, that it was unlikely they would experience any adverse effects as a result of the task, but if they felt disturbed in any way they could contact the researcher, or seek help at the Student Counselling Centre.

The research was not supported by any funding or organization

### **2.4.2 Data collection**

In two separate group administrations, 5 TAT cards, the M-TAT for the control group, and the e-TAT for the experimental group, were presented to participants on a screen using an overhead projector. Based on the method employed by Bailey and Green (1977), participants were given five minutes to write a story about each card. Each

participant was supplied with an answer booklet that had a unique number or code. For purposes of confidentiality they were not asked to write their names, but supplied basic demographic details including race and age. An informed consent declaration was presented on the covering page of the answer booklet and participants were free to leave before, or during the test, with no explanation required.

Identical instructions, based on Murray's (1943) suggestions (see p. 14), were given to both groups. Participants were prompted to write a narrative that contained a beginning, middle and an end. They were also encouraged to furnish the characters with emotion where possible.

### **2.4.3 Scoring Procedures**

Two separate scoring procedures were used. A word count was used to measure the length of the narratives and Gerver's *scoring level of response* was employed to assess the quality of 40 protocols. These scoring systems are discussed in detail below.

#### **2.4.3.1 Word Count**

Each participant was requested to produce 5 narratives corresponding to the 5 different pictures they were presented with. Each narrative was counted and given a *word count* score. All words, including abbreviations, numbers and titles of the narratives were counted. Hyphenated names, such as Mary-Sue, were treated as 1 word. Scripts were excluded if students completed less than 3 narratives, did not supply demographic details, or were obviously not taking the task seriously.

#### **2.4.3.2 Productivity rating**

A random sample of the participants' responses to the technique (both M-TAT and e-TAT) received a score based on Gerver's *scoring level of response* (as cited in Coleman, 1947). It is essentially a productivity score based on the level of interpretation and quality of narrative. An outline of the scoring criteria follows:

### *First Level*

Level one is the first and lowest level of response in which the participant does not respond at all, or writes something completely unrelated to the card presented. The scoring weight is 1.

### *Second Level*

A level two response consists of a static listing of the items or figures in the card. An example might be "*this boy is young and has brown hair and is wearing a shirt*". The scoring weight is 2.

### *Third Level*

In a third level response the participant should describe some form of action, such as looking, kissing, running, hiding, and finding. A level three response calls for more action than a level two response which is static. An example of a third level response could be "*The man is leaving and the girl is holding him back*" or "*The girl is going to school, the man is working on the farm and the lady is watching them*".

### *Fourth Level*

A fourth level response requires an element of interpretation. A participant might interpret what a character's job is or what religion they are. Interpretation for a fourth level response also includes emotional interpretation whereby, the character is assigned a feeling state. A third kind of interpretation is the assigning of character traits, saying for example if the character is mean, kind, evil or good. A response is also given a weighted score of four if there is an example of someone thinking. An example might be "*He is an evil man and is trying to leave. She is scared, and guilty and thinks it is her fault*".

### *Fifth Level Response*

There are two criteria necessary that to achieve a fifth level response which has a weighted score of five. The narrative must contain a plot with a clear beginning, middle and end (including an explicitly stated outcome), and the author must say how one character feels about the other, or how one character feels about his/her situation. Even if

there is no visible other in the card, an external person or opinion could be implied or referred to.

All word counts and productivity ratings were performed by the author. It was expected that word counts would be a predictor of quality in that the longer a story was, the deeper and more emotion driven it would be. Data analysis and results are discussed in the following section (will put section number when whole document is formatted).

## Chapter 3 Results

### 3.1 Overview

This research aimed to investigate whether or not the version of TAT presented (M-TAT or e-TAT) as well as the race group a person identified with, and the interaction between the two, had an effect on aspects of the TAT narratives of psychology students at the University of KwaZulu-Natal (UKZN). The effects investigated include, a *mean word count* of the narrative, as well as a *mean productivity score*, derived from Gerver's *scoring level of response* (as cited in Coleman, 1947, and outlined in section 2.4.3.2). Two 2X2 factorial ANOVAS were conducted, one for the dependent variable *mean word count* ( $n = 207$ ) and one for the other dependent variable, *mean productivity score* ( $n = 40$ ). Thirteen cases were excluded. Ten cases were excluded because they wrote less than three narratives, two cases were excluded due to a lack of demographic details, and one case was left out of the analysis because the student drew pictures, wrote inappropriate narratives and obviously had not taken the task seriously.

**Table 1.**  
Sample Demographics

Demographics	Control group		Experimental group	
	Number	% of group	Number	% of group
1. black	28	68.29	127	76.51
2. white	13	31.71	39	23.49
3. male	10	24.39	49	29.52
4. female	31	75.61	117	70.48

*Note.* The control group received the M-TAT and experimental group received the e-TAT.

### 3.2 The effect of TAT and Race Group on Word Count

To test whether or not the version of TAT administered, as well as the race group the participant identified with, has an effect on the mean number of words produced across 5 stories, a 2 (*TAT* – ‘M-TAT’ versus ‘e-TAT’) X 2 (*Race* – white versus black) factorial ANOVA was conducted. Assumptions of normality and homogeneity of variance were upheld.

Descriptive statistics for this analysis, and a summary of the results, are given in Tables 2 and 3 below. The results revealed no significant main effect for *version of TAT*,  $F(1, 203) = 1.61, p = 0.206$ , indicating that whether the M-TAT or the e-TAT was administered, did not significantly affect the number of words produced. Similarly there was no statistically significant main effect for *race*,  $F(1, 203) = 2.68, p = 0.103$ , therefore the race group that the participant identified with did not effect the mean number of words produced across the two TAT levels. No significant interaction between *version of TAT* administered and *race* was found,  $F(1, 203) = 0.129, p = 0.720$ .

**Table 2.**  
Descriptive Statistics for Mean Number of Words Produced

Version of TAT	Race					
	white			black		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
1. M-TAT	13	80.17	18.49	28	74.43	21.16
2. e-TAT	39	87.49	25.52	127	78.52	23.80

However, while the results were not statistically significant, on examination of the descriptive statistics (Table 2), mean differences reflect positively in favour of the e-TAT, where both black ( $M = 78.52$ ) and white ( $M = 87.49$ ) participants produced more words on the e-TAT than black ( $M = 74.43$ ) and white ( $M = 80.17$ ) participants did on the M-TAT. Furthermore, according to the descriptive statistics (Table 2), black participants produced slightly shorter stories on both the M-TAT ( $M = 74.43$ ) and e-TAT ( $M =$

78.52), than the white participants on the M-TAT ( $M = 80.17$ ) and the e-TAT ( $M = 87.49$ ).

**Table 3.**  
Analysis of Variance for Mean Number of Words Produced

<i>Source</i>	<i>df</i>	<i>F</i>	$\eta^2$	<i>p</i>
Version of TAT	1	1.61	<.01	.206
Race	1	2.68	.01	.103
TATX RACE	1	0.129	<.01	.720
Error	203	(553.2)		

*Note.* Values in parentheses represent mean square errors.

\* $p < .05$ . \*\* $p < .01$ .

While no significant interaction was found, a reflection on the mean differences does reflect favourably in the direction of the experimental hypotheses. A slight increase in means can be seen for black students on the e-TAT as opposed to the M-TAT, and a similarly small difference can be seen on the M-TAT, where black students produced marginally shorter narratives than the white students in that sample.

### 3.3 The effect of TAT and race group on mean quality score

Descriptive statistics and a summary of results for the second factorial ANOVA, where the dependent variable is *mean productivity score*, are presented in Tables 5 and 6 below.

**Table 4.**  
Sample Demographics

Demographics	Control group		Experimental group	
	Number	% of group	Number	% of group
1. black	10	50	10	50
2. white	10	50	10	50
3. male	6	30	7	35
4. female	14	70	13	65

*Note.* The control group received the M-TAT and experimental group received the e-TAT.

Assumptions of normality and homogeneity of variance were once again upheld. The analysis yielded a significant main effect for the *version of TAT*,  $F(1, 36) = 5.20$ ,  $p = 0.029$ , where participants on the e-TAT scored significantly higher ( $M = 4.07$ ) than participants on the M-TAT ( $M = 3.64$ ). A significant main effect was also found for *race*,  $F(1, 36) = 9.79$ ,  $p = 0.003$ , where white participants scored significantly higher ( $M = 4.15$ ) (irrespective of what test was administered), than black participants ( $M = 3.56$ ).

**Table 5.**  
Descriptive Statistics for Mean Quality Score

Version of TAT	Race					
	white			Black		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
1. M-TAT	10	4.16	0.40	10	3.12	0.61
2. e-TAT	10	4.14	0.58	10	4.00	0.75

**Table 6.**  
Analysis of Variance for Mean Quality Score

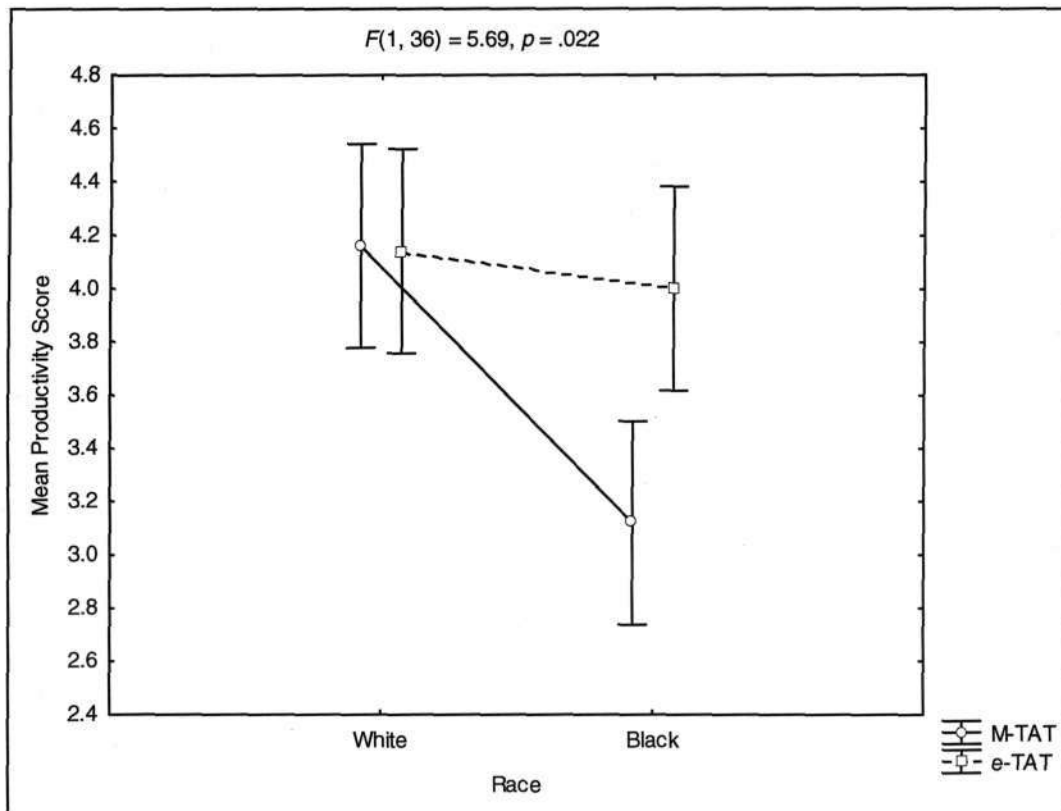
<i>Source</i>	<i>df</i>	<i>F</i>	$\eta^2$	<i>p</i>
Version of TAT	1	5.20*	.09	.029
Race	1	9.79**	.17	.003
TATX RACE	1	5.69*	.10	.022
Error	36	(0.3557)		

*Note.* Values in parentheses represent mean square errors.

\* $p < .05$ . \*\* $p < .01$ .

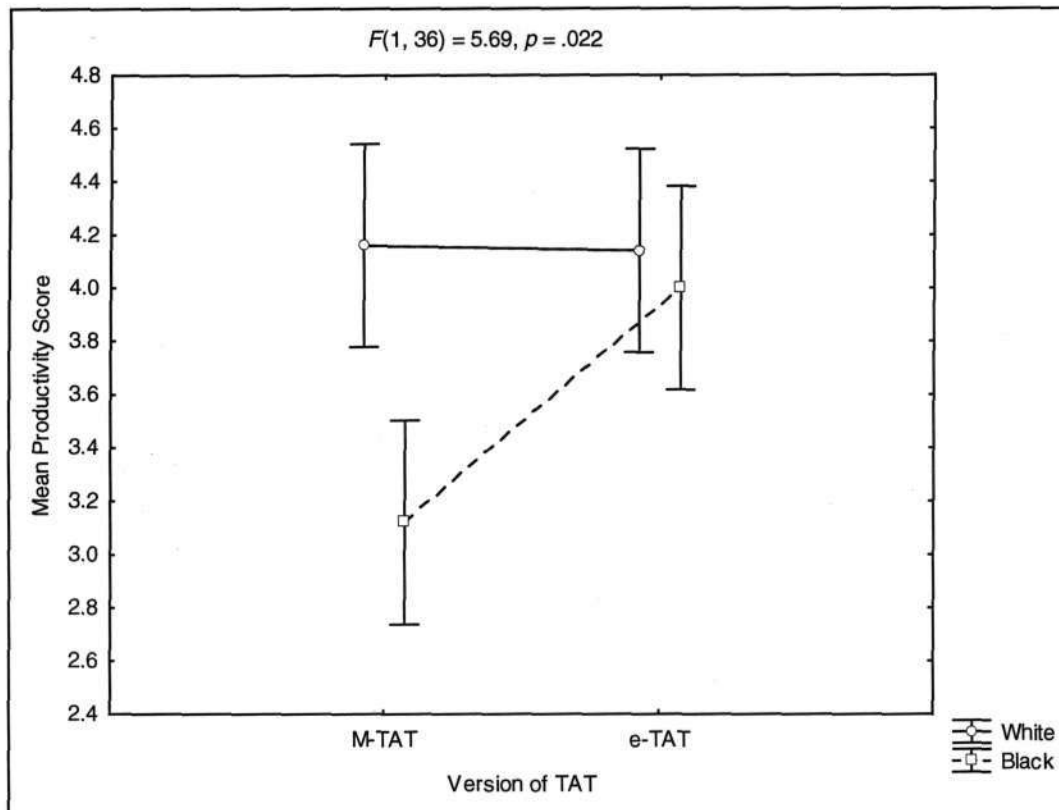
The interaction effect prompted a sequence of post-hoc analyses. These analyses aimed to investigate whether or not the difference between the version of TAT administered was significant for both race groups (black and white), as well as whether or not the difference between race groups was significant for the M-TAT condition. All these analyses were conducted using simple one-way ANOVAs. In order to increase the power of these analyses, the *F*-values were recalculated by dividing the mean square effect of the one-way ANOVA by the mean square error of the factorial ANOVA. The *p*-values were recalculated with the probability calculator function of STATISTICA, using the new *F*-value, and the degrees of freedom effect from the one-way ANOVA and the degrees of freedom error from the factorial ANOVA.

Examination of the first cell mean plot (Figure 1) indicated that it was unlikely that a significant difference existed between white participants on the e-TAT and white participants on the M-TAT, and thus this analysis was not performed. Similarly, on inspection of the second cell mean plot (Figure 2), it appeared unlikely that there would be a statistically significant difference in scores between white and black participants on the e-TAT, and thus this analysis was not conducted.



**Figure 1.** Means plot of the interaction between Race and version of TAT on mean productivity

For both one way ANOVAs, the assumptions of normality and homogeneity of variance were upheld. The one way ANOVA for black participants on the M-TAT versus the e-TAT condition yielded highly significant results,  $F(1, 36) = 10.886, p = 0.002, \eta^2 = 0.32$ . Black participants scored significantly higher on the e-TAT ( $M = 4.00$ ) than the M-TAT ( $M = 3.12$ ), where 32% of the total variability in test scores is accounted for by the version of TAT administered. Similarly, the result of the analysis on white participants versus black participants on the M-TAT was highly significant at the 1% level,  $F(1, 36) = 15.20, p < 0.000$ .



**Figure 2.** Means plot of the interaction between Version of TAT and Race on mean productivity score

White participants scored significantly higher ( $M = 4.16$ ) than black participants ( $M = 3.12$ ). Taking  $\eta^2$  as the measure of effect size, the race group of the participants accounted for 53% of the total variability in the productivity scores ( $\eta^2 = 0.53$ ).

### 3.4 In Summary

The results can be summarised as follows. No significant main or interaction effects were found for the dependent variable *word count*, however, an examination of mean differences reflects favourably upon the e-TAT. Statistically significant main and interaction effects for the dependent variable, *quality*, indicated that (1) across both race groups, participants scored higher on the e-TAT, (2) black participants attained significantly poorer results on the M-TAT than white participants on the M-TAT, and (3) black participants produced significantly higher quality protocols on the e-TAT than

black participants on the M-TAT. In the following section, the implications of the results that emerged from the statistical analyses are discussed and contextualised.

## Chapter 4 Discussion

### 4.1 Theoretical Overview

This study set out to explore whether a modified TAT generated better narrative protocols than those on the Murray TAT. If so, this would lend support to an argument for the development of a new culturally and racially inclusive South African TAT. In broad terms, this study set out to investigate whether there is a need to develop a new, culturally *and* racially congruent TAT for South Africans. Culture and race were looked at separately, as it is argued that they are not simply synonymous in the complex and dynamic reality of urban South Africa. The term culture has many different definitions, but for the purposes of this study, it is understood as something *learned* rather than something inherited, such as the colour of a person's skin (Goodenough, 1994).

As a very old definition suggests, culture, in this thesis, is understood to include, knowledge, art, morals and all other capabilities that humans acquire to become members of a society (Tylor, 1871). Furthermore, culture is learned in order for people to *interact* with others in their society (Thornton, 1988) and cultural realities are actively created by people, and shaped by their environment (Goodenough, 1994). The detailed theoretical relationships between culture and race are beyond the scope of this work, however the terms are separated in a stance that echoes Boonzaier and Sharp's (1988) comment that "the classic error in South Africa is to presume that if people are identified, or identify themselves as 'Zulu', this label marks the essence of their being, and says all there is to say about them" (p. 15). This view is extended, to argue that identifying oneself as black or white does not say "all there is to say" about a person.

In addition, this study takes as axiomatic, that people have not one, but multiple cultural identities (Swartz, 1992). It is possible therefore, to have a family cultural identity, that includes language, symbols and behaviours congruent with ones heritage. It is similarly legitimate however, to *simultaneously* have something akin to a 'university student' cultural identity. While it is somewhat simplified, this study treats 'culture' as something

South African rather something than racial<sup>8</sup>. It was assumed that a 'South African culture' in contrast to a 'European culture' for example, is a framework potentially shared by students of all race groups in the sample.

The research aims were derived from an interest in both cultural and racial issues with regard to the Murray TAT, a technique neither developed nor normed in South Africa.

Firstly, with regard to culture, this research aimed to investigate the differences between the effects of a revised TAT (e-TAT) and the original Murray TAT, on a sample of racially mixed first year psychology students at UKZN. In predicting the potentially different effects of the two TATs, the notion of a shared cultural reality for young urban South Africans that transcends race was assumed. These aims of the research were derived from the following questions: (1) Do South African protocols on an experimental TAT, that is racially inclusive, differ significantly in quality and length from South African protocols on the original Murray TAT cards? Do black *and* white South Africans, resonate more strongly, as expressed in the dependent variables word count and quality rating of this study, with stimuli that include black characters (and are hence arguably more culturally congruent in South Africa) than with the M-TAT, that consists of white people only?

Secondly, with regard to race, the study aimed to investigate whether a difference in scores between two different TATs was dependent on the race group that the participant self-identified with. The assertion by Bailey and Green (1977), that "TAT material which is matched to the racial group being tested is valuable in enhancing subject responses" (p. 30) prompted the following questions: (1) Does the Murray TAT inhibit the responses of black South Africans? And (2) Would a TAT that included depictions of black people enhance the quality and length of TAT protocols for black South Africans?

The research hypotheses can hence be summarised as follows:

---

<sup>8</sup> This is certainly not to say that people *don't* have cultural identities that are shaped by their ethnicity or heritage, but that being 'South African' is one possible aspect of their culture, or one cultural identity among many.

- The first experimental hypothesis was that there would be a significant difference in story length, with all participants producing longer stories on the e-TAT than on the M-TAT.
- The second experimental hypothesis was that there would be a significant difference in story length, with black participants producing longer stories on the e-TAT than on the M-TAT.
- The third experimental hypothesis was that there would be a significant difference in story length, with black participants producing shorter stories on the M-TAT than white participants on the M-TAT.
- The fourth experimental hypothesis was that there would be significantly greater quality of narrative for all participants on the e-TAT than on the M-TAT
- The fifth experimental hypothesis was that there would be significantly greater quality of narrative for black participants on the e-TAT than on the M-TAT.
- The sixth experimental hypothesis was that there would be significantly greater quality of narrative for white participants on the M-TAT than black participants on the M-TAT.

## **4.2 Implications of Results:**

### **4.2.1 Results for word count**

Contrary to the first experimental hypothesis that all participants would produce significantly longer narratives on the e-TAT than the M-TAT, no statistically significant main effect was found for the version of TAT administered. These results reflect Narron's (2005) findings that participants did not write longer stories (more words) on a more modern TAT than they did on the M-TAT. What Narron (2005) did argue however, was that content of the narratives on the M-TAT might be more reflective of the era in which the original cards were produced, rather than true projective engagement with the stimuli. The use of word counts as a measure of projection is more critically discussed in the section 'Quantity is not quality' below (section number).

A main effect for race was not expected, in that this study did not aim to determine whether there was a significant difference in story length between races, *regardless* of the TAT administered. It was not expected for either race group to write longer or shorter narratives across both tests, and accordingly, there was no main effect for race group found when looking at mean word counts.

The second experimental hypothesis, that black participants would produce longer narratives on a TAT that was racially inclusive, than on the M-TAT, was similarly discounted by the lack of significant interaction found (Table 2.). With regard to the third experimental hypothesis, black participants did not write significantly shorter stories on the M-TAT than the white participants on the M-TAT. The data from this study did not prove that racially incongruent stimuli would result in more defended and therefore shorter narratives (Bullard, 2005). An alternative interpretation however, is that the e-TAT did not sufficiently encapsulate the personalities or life experiences of the participants it aimed to assess.

Based on these findings alone, questions arise as to the utility of using racially inclusive and more modern TAT material in South Africa, or question the value, and applicability of the constructed e-TAT. The results for the dependent variable, *word count*, however, are juxtaposed with the results of a sub-sample, who received a productivity rating, intended to measure the quality of their narratives.

#### **4.2.2 Results for productivity**

In contrast to the results on mean word count, highly significant results emerged for the sample who received productivity rating scores. There was a significant main effect for both race and version of TAT, as well as a significant interaction. Howell (2004), asserts that while many people argue that main effects should be ignored in the presence of a significant interaction, it is not reasonable to automatically discount main effects if these effects make sense. Furthermore, main effects are most often theoretically meaningful

when they are ordinal (the ordinal interaction can be seen in Figure 1. and Figure 2.) and therefore it appears appropriate to interpret them (Howell, 2004).

As predicted in the fourth experimental hypothesis, the statistically significant main effect for the version of TAT administered, indicates that when the scores are averaged across both levels of race, participants wrote higher quality narratives on the e-TAT than they did on the M-TAT. This result could be interpreted as support for the idea that all South Africans are resonant with, and therefore inspired to tell more emotion based narratives, when faced with stimuli that are representative of the racially mixed environment in South Africa. However, recalling Howell's (2004) caution in interpreting main effects, the possibility remains that this improvement in all South African's scores could be unduly affected by a big improvement in black participants' scores. Indeed, as discussed below, it can be seen that through analysis of white participants' scores on the M-TAT versus the e-TAT, this 'cultural' hypothesis is not supported by the results.

The statistically significant main effect for race, that is less meaningful in terms of the research question, but nonetheless important to mention, indicates that when the scores are averaged across both the e-TAT and M-TAT, white participants produce higher quality narratives than black participants. Owing to the minimum level of education in the sample being a first year university level, it is difficult to attribute this difference in scores solely to language diversity (English fluency). Nevertheless, it should be noted that, while language proficiencies were not requested, due to a substantial proportion of the black sample being African, there is a high probability that some students were second language English speakers. It is beyond the scope of this study to discuss the possible effects of language on the quality of narratives in detail, yet a basic awareness of the potentially restraining effect of using one's second language in 'creative writing', is important.

The statistically significant interaction effect suggests that the difference in mean scores between the race groups (namely black versus white) depends on the version of TAT administered (namely e-TAT or M-TAT). Similarly, it can be said that the difference between mean scores on the two versions of TAT, depends on the race group of the

participant. On examination of the post hoc analyses, some meaningful results emerged, confirming the experimental hypotheses. These are discussed more fully below.

On the M-TAT, black participants wrote significantly lower quality narratives than the white participants did. It could thus be argued that there is bias inherent in most contemporary practice of the M-TAT in South Africa. Using the M-TAT with black South Africans may be limiting, in that an interpreter might understand the less dramatic, and emotion based narratives as defensiveness from the test-taker. Moreover, as Bullard (2005) asserted, in the history of South African politics and psychology, results such as these have been exploited by interpreting them in isolation. Colonial psychiatry is notorious for using black people's poor results on 'Western' tests to 'scientifically claim' their inferior intellect (Ainsworth, 1959, Bullard, 2005). It is important to interpret the poorer quality of narrative for black participants on the M-TAT, in light of the significant improvement of black participants on the e-TAT.

Comparing black participants' narratives on the M-TAT to the narratives of black participants on the e-TAT, it can be seen that participants on the e-TAT wrote significantly higher quality stories. This result could be interpreted as corresponding with Hock's (1999) suggestion that people are most likely to engage in dramatic story telling and potential projection when they identify with a character in the card. Accepting this interpretation lends support to the idea that identification and projection are increased when there is congruence between the respondent and the TAT stimulus.

The results for narrative quality suggests that white participants attained scores on the e-TAT that were very similar to those attained by white participants on the M-TAT. On inspection of the cell mean plots (Figure 1. and Figure 2.) almost no difference can be seen. This is in contrast to the idea that the M-TAT is equally limiting for white participants as it is for black participants. This result casts doubt on the hypothesis, that in addition to racial bias against black respondents, the M-TAT is outdated for many white South Africans as well (de Bruin, 2001). Similarly the result does not support the notion that 'experience near' or more modern stimuli are required for identification and higher quality narratives (Narron, 2005) and instead shifts the focus more towards racial

congruence rather than 'cultural' congruence. Recalling that Morgan and Murray felt it was important that the participant should easily recognise a character in every card (Hock, 1999), these results suggest that white participants are as familiar with white characters of the 1930's media as they are with the more modern characters depicted in the e-TAT.

### 4.3 'Quantity is not quality'

An unanticipated outcome emerged in that no results for the dependent variable *word count* were statistically significant, but many analyses for the dependent variable *quality* were. While it was expected that a longer narrative would increase the likelihood of a deeper, more emotionally complex plot, this does not seem to have been the case for the discussed sample. While Narron (2005) equated the number of words used for a narrative with stimulus engagement, this does not appear to hold for this sample. Participants on the e-TAT managed to write significantly better quality narratives, without using more words than participants on the M-TAT, who scored significantly lower on the quality rating. Reflecting on the criteria for points on the quality rating (see section 2.4.4), it can be deduced that a small sub-sample ( $n=10$ ) of black participants on the e-TAT used more emotion words, used more interpretation and described plots with outcomes, more often than, but using a similar number of words to, black participants on the M-TAT. The same can be said for white participants' greater quality scores as compared to black participants poorer quality scores on the M-TAT.

A similarly unexpected yet different discrepancy arose in Bailey and Green's (1977) study on TAT variations in an American sample. While a black TAT enhanced participants' subjective ratings of how much the technique facilitated story production, no greater quality of content was expressed by black people on the black TAT, than on the Murray TAT. In other words, subjective perceptions of the black TAT increased, but the quality of content did not.

With regard to the e-TAT versus M-TAT, a possible explanation that cannot be ignored, is that neither *quality* nor *quantity* are correlates of identification or projection. It is possible that both dependent variables in this study are independent measures, each of something other than projection. Recall one of Morgan and Murray's (in Hock, 1999) earliest studies on the TAT when they discovered that narrative responses were most inspired by four main sources. Apart from the test taker's conscious and unconscious fantasies, in addition, stories were often reproductions of popular fiction; lived experiences told by friends or relatives; as well as events in the participant's own life (Morgan and Murray in Hock, 1999). Moreover, Narron (2005) argues that because the Murray cards are so outdated, they are more likely to elicit stories about history, than true projection. In using the word count as a dependent variable, narratives that may have been reflecting the historical period, or reproducing popular fiction and lived experiences rather than projection, would not have been identified. Similarly, in giving a narrative a quality rating score in isolation from discussion with or knowledge of the writer, the likelihood of determining true projection from superficial accounts is unlikely.

The two measures 'quantity' and 'quality' appear to be independent where quantity is not a reliable predictor of quality. The refining of valid dependent variables is important and should be seriously considered in future research.

#### **4.4 The use of aggregates**

A difficult decision arose during the statistical analyses of the data. While the research design originally intended to measure the *mean* scores for both quality ratings and word counts using factorial ANOVAs, there was a question around the potential use of MANOVA or regression analysis to analyse differences on each individual card. It could be argued, that due to a lack of internal consistency (Cramer, 1999) in the TAT type techniques, the individual pictures were unlikely to form a scale, and consequently the average score would be a less meaningful compression of the results. Unlike a cluster of items on a personality scale that are all intended to measure the same personality trait, it

is well known that different cards in the TAT were designed to elicit different themes and to assess different areas of psychological functioning (Cramer, 1999).

As discussed in the literature review (section 3.1.2), Lilienfeld et al. (2000) argue against the use of averages in the overall interpretation of TAT results because of the lack of internal consistency between the stimuli. While Lilienfeld et al. assert the average TAT score is meaningless, Hibbard (2003), in reply to this argument, refers to the meaningful use of aggregates (even in the absence of internal consistency), throughout the discipline of clinical psychology.

It therefore seemed reasonable to retain the original experimental design and use the average scores of participants. Furthermore, in comparing participants' scores, the average score was not meaningful in itself, but meaningful in the differences between averages that were predicted to emerge. This is not to say it would not be useful to investigate each stimulus card, but that it is beyond the scope of this pilot study to do so.

#### **4.5 Limitations and Suggestions for Future Research**

The limitations of this study include issues around the sample size and structure. While it was expected that the two lecture slots would be attended by approximately half the psychology first year group each, it turned out that the experimental and control groups were vastly different in size. The calculations in factorial ANOVA become considerably more complicated in the event of unequal sample sizes. However, provided that the assumptions of normality and homogeneity of variance are met, factorial ANOVA is robust in this respect (Howell, 2004). In relation to sample sizes, it is also a limitation of the research that only a smaller randomly selected sub-sample of the participants could be rated on the quality of narrative. This was due to limited scope and resources. It is suggested that in future research, the whole sample is scored on both quality and word count to allow for more valuable comparisons.

Due to the current study being a pilot, and because of limited resources, only 5 cards were adapted and administered. Future research may benefit from constructing and using up to ten cards, which could arguably give respondents time to get used to the story writing process. Once participants are familiar with the task, they may start to feel more comfortable, and longer narratives could emerge. The converse however may also be true in that participants may tire of the task after a few cards or be overwhelmed by all the writing expected of them.

Owing to the apparent independence of word count and quality rating for this study, an extension of this research might include comparing full clinical analyses of all stories, interpreted blind by expert raters. This kind of methodology also has the potential to overcome the possible confounding factor, that some narratives high in word count and 'quality' (by Gerver's scoring level of response) may in fact not be representative of projective engagement with the stimuli. As previously discussed, theorists have asserted that respondents on a TAT type technique might produce stories reflective of the historic period of the M-TAT (Narron, 2005), or reproductions of recent events and popular fiction (Morgan & Murray, in Hock, 1999). While it could be argued that there is even interpretive value in the way respondents express reproductions of history or popular fiction, as Murray (1951) stated the true value of the TAT "is its capacity to reveal things the subject is unwilling to tell or unable to tell..." (p. 577).

The use of mean scores of the participants in the sample is not understood as a limitation, however it could be proceeded by an investigation of the potentially different effects or resonance value of different cards. If there was significant evidence suggesting the need for a new South African TAT, then individual cards would need to be independently analysed. Cards that were not found to evoke higher quality narratives could be improved upon or replaced. In addition, it might be useful for a panel of expert TAT clinicians to evaluate the adapted version before it is administered to the sample of participants.

## 4.6 In Conclusion

On the basis of the study's findings, it is concluded that there seems to be some improvement in the quality of TAT narratives for black participants on a racially inclusive adaptation of the TAT cards. Further it was shown that while the e-TAT does not significantly improve white participants' quality of narrative response, there is no sign of a decrease in the quality or length of their stories. The finding that participants' word counts did not significantly increase on the presentation of the e-TAT does not have to be interpreted as evidence against the construction of a TAT for South Africans, but might indicate that the *length* of protocols is not the only, or even most appropriate, measure of projective engagement. There appears to be more evidence supporting the need for racially congruent, rather than culturally congruent TAT material, but in light of the prejudice perpetuated by racially exclusive tests in the history of South African psychometrics, these results are presented with caution.

The results that emerged through this study are favourable but not yet conclusive. It is hoped that the further research needed to develop, ensure and maintain fair testing practice in our diverse country will be continued.

## References

### References: Primary Sources

Murray, H.A. (1943). *Thematic Apperception Test manual*. Cambridge, MA: Harvard University Press.

### References: Related publications

Ainsworth, M. (1959). Thematic Apperception Test: Thompson modification. In O.K. Buros (Ed.), *The Fifth Mental Measurements Yearbook*. (pp.310-315). New Jersey: The Gryphon Press.

Bailey, B., & Green, J. (1977). Thematic apperception test stimulus material. *Journal of Personality Assessment*, 41, 25 – 30.

Boonzaier, E. & Sharp, J. (1988). *South African Keywords*. Cape Town: Philip

Bracken, P., & Thomas, P. (2001). Post psychiatry: New direction for mental health. *British Medical Journal*, 322, 724-727.

Bullard, A. (2005). The critical impact of Franz Fanon and Henri Collomb: Race, gender, and personality testing of North and West Africans. *Journal of the History of the Behavioral Sciences*, 41, 225 – 248.

Coleman, W. (1947). The thematic apperception test. I. Effect of recent experience. II. Some qualitative observations. *Journal of Clinical Psychology*, 3, 257 – 264.

Colman, A. (2001). *Dictionary of psychology*. Oxford: Oxford University Press.

- Corey, G. (2005). *Theory and practice of counseling and psychotherapy*. United States: Brooks/Cole.
- Cramer, P. (1999). Future directions for the Thematic Apperception Test. *Journal of Personality Assessment*, 72, 74-92.
- De Bruin, G. (2001). Personality assessment. In C. Foxcroft and G. Roodt (Eds.), *Introduction to psychological assessment in the South African context* (pp.227-248). Cape Town: Oxford University press southern Africa.
- Ehrenreich, (1990). Effect of social class of subjects on normative responses to TAT cards. *Journal of Clinical Psychology*, 46, 467 – 471.
- Frank, L. (1971). Projective methods for the study of personality. In L.D. Goodstein & R. I. Lanyon (Eds.), *Readings in personality Assessment*. (pp. 132-147) New York: John Wiley & Sons, Inc.
- Gaines, A. (1992). Ethnopsychiatry: The cultural construction of psychiatries. In A. Gaines (Ed.), *Ethnopsychiatry*, (pp. 3-51). New York: State University of New York Press.
- Goldberg, P.A. (1971). A review of the sentence completion methods in personality assessment. In L.D. Goodstein & R. I. Lanyon (Eds.), *Readings in Personality Assessment* (pp. 147-184). New York: John Wiley & Sons, Inc.
- Goodenough, W. (1994). Towards a working theory of culture. In R. Borofsky (Ed.) *Assessing Cultural Anthropology* (pp. 262-273). New York: Mcgraw Hill.
- Hibbard, S., Tang, C. & Latko, R. (2000). Differential validity of the defense mechanism manual for the TAT between Asian Americans and whites. *Journal of Personality Assessment*, 75, 351 – 372.

- Hibbard, S. (2003). A critique of Lilienfeld et al.'s (2000) "The scientific status of projective techniques". *Journal of Personality Assessment*, 80, 260-271.
- Hock, R. (1999). Psychotherapy. In R. Hock (Eds.), *Forty studies that changed psychology. Explorations into the history of psychological research (3<sup>rd</sup> ed)* (pp. 254-275). New Jersey: Prentice-Hall.
- Hofer, J. & Chasiotis, A. (2004). Methodological considerations of applying a TAT-type picture-story test in cross cultural research. A comparison of German and Zambian adolescents. *Journal of Cross Cultural Psychology*, 35, 224-241.
- Howell, D. C. *Fundamental Statistics for the Behavioral Sciences (5<sup>th</sup> ed.)*. U.S.: Brooks/Cole
- Lilienfeld, S.O., Wood, J.M. & Garb, H.N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, 1, 27-65.
- Louw, D. & Edwards, D.J.A. (1997). Motivation and Emotion. In D. Louw & D.J.A. Edwards (Eds.), *Psychology. An introduction for students in Southern Africa (2<sup>nd</sup> ed)*. Johannesburg: Heinemann.
- Makhale-Mahlangu, P. (1996). *Reflections on trauma counselling methods*. Retrieved May 17, 2008, from The centre for the study of violence and reconciliation Web site:  
<http://www.csvr.org.za/wits/articles/artpales.htm>
- Mkhize, N. (2004). Psychology: An African Perspective. In D. Hook (Ed.), *Critical Psychology* (pp. 24-52). Cape Town: UCT Press.

- Monopoli, J. & Alworth, L. (2000). The use of the Thematic Apperception Test in the study of Native American Psychological Characteristics: A Review and archival study of Navaho Men. *Genetic, Social, and General Psychology Monographs*, 126, 43-78.
- Murray, H. (1951). Uses of the Thematic Apperception Test. *American Journal of Psychiatry*, 107, 577-581.
- Naidu, B. (2008, January 20). Pardon my language. *University of Johannesburg News*. Retrieved May 17, 2008, from <http://www.uj.ac.za/latestnews>.
- Narron, M. (2005). *Updating the TAT: A photographic revision of the thematic apperception test*. USA: Proquest information and learning company.
- Rabin, A.I. (2001). Projective Techniques at mid-century: A retrospective review of an Introduction to Projective techniques by Harold H. Anderson and Gladys L. Anderson. *Journal of personality assessment*, 76, 353-367.
- Retief, A. (1987). Thematic apperception testing across cultures: Tests of selection versus tests of inclusion. *South African Journal of Psychology*, 17, 47 – 54.
- Rossini, E., & Moretti, R. (1997). Thematic Apperception Test (TAT) interpretation: Practice recommendations from a survey of clinical psychology doctoral programs accredited by the American Psychological Association. *Professional Psychology: Research and Practice*, 28, 393-398.
- Rotter, J.B. (1946) Thematic apperception tests: suggestions for administration and interpretation. *Journal of Personality*, 15, 70-92.

- Sikhosana, M.J. (1986). *An investigation into the predictive validity of the Thematic Apperception Test (Zulu) (TAT-Z) as a method for assessing leadership*. Unpublished Masters Thesis, University of Natal, Pietermaritzburg.
- Swartz, L. (1992). Professional ethnopsychiatry in South Africa: the question of relativism. In A. Gaines (Ed.), *Ethnopsychiatry* (pp.225-251). New York: State University of New York Press.
- Thornton, R. (1988). Culture: a contemporary definition. In E. Boonzaier & J. Sharp (Eds.), *South African Keywords* (pp. 17 – 28). Cape Town: Philip.
- Tomkins, S. (1947). *The Thematic Apperception Test. The theory and technique of interpretation*. New York: Grune and Stratton, inc.
- Tylor, E. B. (1871). *The Science of Culture*. London: J Murray.

## Appendix A

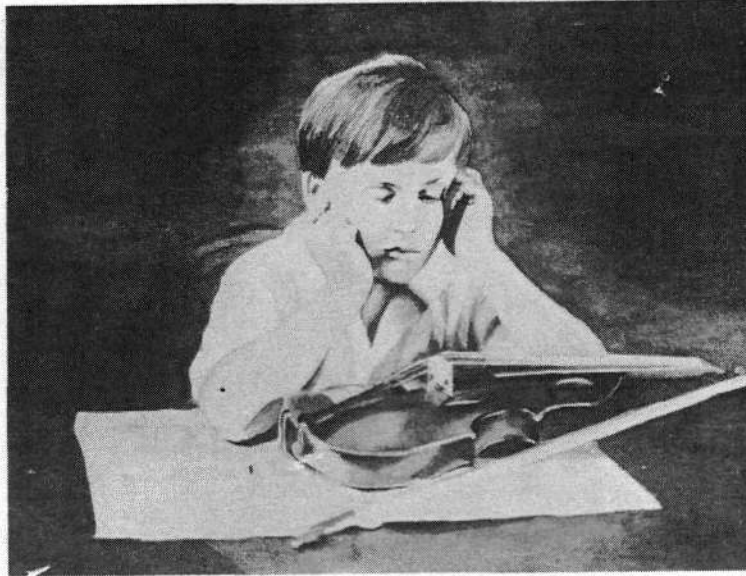


Figure 3. Murray TAT Card 1



Figure 4. Experimental TAT Card 1



**Figure 5.** Murray TAT Card 2



**Figure 6.** Experimental TAT Card 1



**Figure 7.** Murray TAT Card 4



**Figure 8.** Experimental TAT Card 3



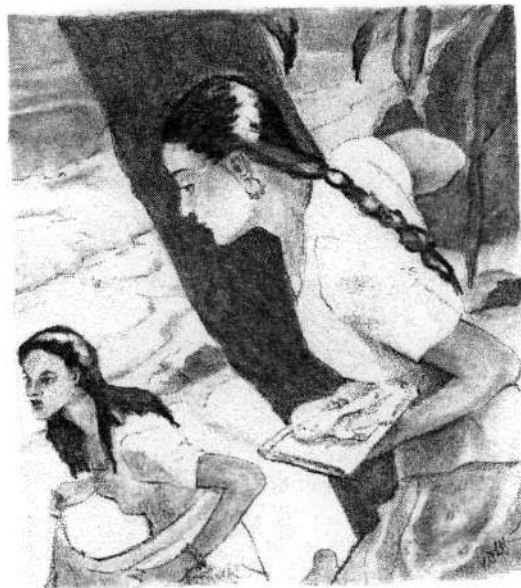
**Figure 9.** Murray TAT Card 6GF



**Figure 10.** Experimental TAT Card 4



**Figure 11.** Murray TAT Card 9GF



**Figure 12.** Experimental TAT Card 5