

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**Investigation of the Humm Wadsworth Temperament Scale:  
Revision, Development and Application**

**This thesis is presented in partial fulfillment of the requirements for the degree**

**of**

**Master of Arts**

**in**

**Psychology**

**at Massey University, New Zealand**

**Kimberley Severinsen**

**2006**

## ABSTRACT

The present investigation examines the psychometric properties of a measure of temperament, the Humm Wadsworth Temperament Scale (Humm). To this end, participants ( $n = 27,245$ ) completed the Humm questionnaire as part of either a recruitment and selection process initiated by a prospective employer, a promotion and development assessment initiated by their current employer, or career guidance advice sought of their own volition. Quantitative theoretical analysis based on Thurstone's method of paired comparisons and conceptual analysis by Humm experts and users were utilised for both the single-loading items for each of the seven components of the Humm, as well as the remaining multi-loading items. Thurstone's method was used to rank order items conceptually from 'best' predictor to 'worst' predictor of a certain component, which in turn were used to identify which items should remain in the Humm and which items should be discarded. The conceptual judgments generated by Humm experts and users, followed by confirmatory factor analysis, were used to increase the validity of the Humm through revising the set of items in the version of the Humm currently in use. The study concludes with a discussion of issues surrounding psychometric test revision, applicability of the Humm to the wider community including culturally diverse populations, as well as suggestions and recommendations for future research in this area.

## **ACKNOWLEDGEMENTS**

I would like to sincerely thank my supervisor, Dr Richard Fletcher for his encouragement, support and assistance throughout the research and thesis process. My appreciation is also extended to the human capital solutions company involved in this project for providing support for a scholarship application, and in particular thanks to my manager and the Research and Development team who have provided me with access to the psychological measurement instrument and data, without which, this thesis would not have been possible. I would also like to express my gratitude to the Tertiary Education Commission for providing me with support in the form of an enterprise scholarship. Finally, I would like to thank my colleagues, friends and family who have given me endless support, encouragement and motivation throughout my Masters study.

It is noted that this project was judged to be low risk. Notification of this was provided to the Massey University Human Ethics Committee.

## TABLE OF CONTENTS

INTRODUCTION .....	1
The Concept of Temperament and Personality .....	2
The Humm Wadsworth Temperament Scale .....	7
The Humm's Seven Components of Temperament .....	19
<i>Normal</i> .....	19
<i>Hustler</i> .....	20
<i>Mover</i> .....	20
<i>Artist</i> .....	21
<i>Politician</i> .....	21
<i>Engineer</i> .....	22
The Humm in relation to other Personality Measurement Instruments .....	22
Evidence for Revision of Psychological Measurement Instruments.....	25
Measurement .....	26
Thurstone's Method of Paired Comparisons .....	28
Aims and Rationale of the Current Study .....	30
METHOD.....	31
Participants.....	32
Apparatus .....	33
Biographical Data Task.....	35
Test Procedure.....	36
Thurstone's Method of Paired Comparisons for Single-loading items.....	37
Expert judgments for Multi-Loading items .....	38
Confirmatory Factor Analysis.....	40
Analysis.....	43
RESULTS .....	46
Analysis of the Current Humm .....	46
<i>Current Normal Single Factor Model</i> .....	47
<i>Current Hustler Single Factor Model</i> .....	47
<i>Current Mover Single Factor Model</i> .....	47
<i>Current Double-Checker Single Factor Model</i> .....	48
<i>Current Artist Single Factor Model</i> .....	48
<i>Current Politician Single Factor Model</i> .....	48
<i>Current Engineer Single Factor Model</i> .....	49
<i>Current Seven-Factor Model</i> .....	49
Development of a Revised Humm .....	50
<i>Revised Normal Single Factor Model</i> .....	51
<i>Revised Hustler Single Factor Model</i> .....	52
<i>Revised Mover Single Factor Model</i> .....	53
<i>Revised Double-Checker Single Factor Model</i> .....	54
<i>Revised Artist Single Factor Model</i> .....	55
<i>Revised Politician Single Factor Model</i> .....	55

<i>Revised Engineer Single Factor Model</i> .....	56
<i>Revised Seven-Factor Model</i> .....	59
Validation of Results.....	61
DISCUSSION .....	63
Limitations .....	67
Future Directions.....	71
CONCLUSION .....	75
BIBLIOGRAPHY .....	80

## LIST OF TABLES

1. Participants level of English at time of completing questionnaire.....	p33
2. A sample of Humm questionnaire items for each component.....	p34
3. Percentage of agreement between Humm experts for multi loading items.....	p39
4. Goodness of Fit indices for current Humm single factor models.....	p49
5. Correlations between components for current Humm seven factor model.....	p50
6. Goodness of Fit indices for revised Humm single factor models.....	p59
7. Correlations between components for revised Humm seven factor model (with statistically selected E).....	p60
8. Goodness of Fit indices for current and revised Humm seven factor models.....	p61
9. Second CFA Goodness of Fit indices for current and revised Humm seven factor models.....	p62

## LIST OF FIGURES

1. The Seven Components and 31 Sub-components of the Humm.....	p18
2. The Big Five Factors and 30 Facets of the NEO-PI-R.....	p24
3. Allocation of Items across the Current and Revised versions of the Humm.....	p45
4. Standardised Regression Weights for the Revised Normal Model.....	p51
5. Standardised Regression Weights for the Revised Hustler Model.....	p52
6. Standardised Regression Weights for the Revised Mover Model.....	p53
7. Standardised Regression Weights for the Revised Double-Checker Model.....	p54
8. Standardised Regression Weights for the Revised Artist Model.....	p55
9. Standardised Regression Weights for the Revised Politician Model.....	p56
10. Standardised Regression Weights for the Revised Engineer Model where items were selected conceptually .....	p57
11. Standardised Regression Weights for the Revised Engineer Model where items were selected statistically .....	p58

**LIST OF APPENDICES**

I. Standardised Correlations for Current Seven-Component Model .....p78  
II. Standardised Correlations for Revised Seven-Component Model.....p79

## **INTRODUCTION**

The purpose of the current study was to evaluate the psychometric properties of a commercial psychological measurement instrument that is utilised to measure temperament, and attempt to improve the measure's statistical validity and interpretation. To investigate this, Thurstone's method of paired comparisons, conceptual judgements by experts and users of the measure, and confirmatory factor analysis methods were utilised in order to place the measure on a sound scientific foundation for further investigation and revision.

The measure chosen for the present study was the Humm Wadsworth Temperament Scale (Humm), developed by Humm and Wadsworth in 1935. The Humm is a psychological measurement instrument that a human capital solutions company utilises to measure temperament and predict behaviour. The current investigation aims to make the Humm more meaningful with respect to the validity and interpretation of the measure, as well as provide conceptual clarity with regards to the questionnaire items that are currently in the Humm. The present research will allow for further development and adaptation of the measure, and increase the ability to generalise the results across the wider community. The information will also aid in the possible adaptation, addition or deletion of current questionnaire items within the Humm. Initially the research will involve reviewing the Humm as a whole, and identifying any biased or ineffective test items through sound statistical analysis.

The concept of temperament and the operational measure, the Humm, are discussed, followed by an exposition on psychological instrument revision and measurement. The

quantitative technique of Thurstone's method of paired comparisons is also discussed, concluding with a summary of the rationale and aims of the current investigation.

### **The Concept of Temperament and Personality**

In the past, the expressions temperament, character, and personality have been used to refer to what is now considered as the term personality (Endler, 1989). However, Endler suggests that temperament refers to the material that personality evolves from, whereas personality is the manner in which a person interacts with themselves and their environment. Whilst there has been much debate about whether personality and temperament are actually one term referring to the same concept (Strelau, 1987; Goldsmith & Campos, 1982), many researchers have continued to use the terms personality and temperament interchangeably (Pervin, 2002; Borkeanu, 2001; Gray, 1973; Sheldon & Stevens, 1942). Furthermore, measures that have been designed to assess either temperament or personality may well have commonalities due to the possibility that they are actually measuring the same variables (Endler, 1989). Thus for the current study, the terms personality and temperament are assumed to be referring to the same concept and are indeed used interchangeably.

Interest in personality and temperament as a predictor of job performance has significantly grown in recent times. So too has the interest in personality measurement. This is in part due a growing number of studies demonstrating that the variables of an individual's personality can predict their future performance across a diverse range of occupations. Additionally, there are an increasing number of measures being made

available to assess temperament (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991). This interest is further supported by other measures of personality that have been successfully used to predict a wide range of occupational performance criteria (Barrick & Mount, 1991, 1993; Hurtz & Donovan, 2000). Hogan and Nicholson (1988) also suggest that an appropriate methodology in many areas of personality and industrial psychology is that of personality assessment. It has now been widely published, acknowledged and accepted that temperament is a predictive measure of performance. When temperament is assessed using well-constructed, valid and reliable measures, the results can be used in personnel selection as a valid predictor of job performance across a variety of occupations (Salgado, 1999).

The term personality includes all factors entering into the make-up of an individual. This can include gender, physical appearance, aptitudes, abilities, talents, disposition and any other factor that may contribute to the whole person and differentiate them from other individuals (Humm & Wadsworth, 1935). Pervin, Cervone and John (2005) define personality as “those characteristics of the person that account for consistent patterns of feeling, thinking and behaving” (p.6). Temperament on the other hand, is used to designate those factors of personality that contribute to disposition, social reactions, emotional tone and attitudes. Temperament determines how a person will behave in a particular situation and how an individual will use their personal resources. In this instance reference is made to behaviour that is based on habits, feelings, attitudes and emotions, rather than behaviour based on purely rational grounds. Temperament is the non-rational and impulsive aspect of a person; for example, a person may possess a temperament style that sees them automatically taking a logical and unemotional stance. In fact, a person might exhibit this

style to such a degree so as to maintain a very unemotional attitude even though many people would suggest that an emotional response would be the “appropriate” response in a particular situation. In other words, even a rational style can be exhibited to an irrational degree. Some people will be naturally more inclined to exhibit “uncontrolled” temperament behaviour, leading to another important point on temperament, the issue of appropriateness. There is neither a right nor wrong temperament style. Being “strong” on one component or characteristic may not necessarily be better than being “weak”. What is inappropriate, when referring to the workplace, is the fit between the person’s temperament style, the job task at hand and the workplace environment. Temperament components are neutral and do not naturally carry positive or negative associations until placed within a context. An individual’s balance of strengths and weaknesses will also vary from environment to environment. People possess all of the temperament characteristics to a certain degree. However, it is the relative degrees and blend of characteristics that create each person’s individuality. Temperament is also quite enduring. Whilst significant life experience will change and alter a person, as will time and maturity, much of the impulsive and non-rational part of a person (their temperament) will remain the same. The question remains, if all people possess all characteristics of temperament, but to differing degrees, how are these characteristics measured? As temperament characteristics are displayed by people to varying degrees, it is possible that the differences and similarities can be compared and therefore can be measured.

Most people will generally display an “average” amount of any given temperament characteristic in comparison to the rest of the population. However, on closer inspection any given person can also fall outside what is considered displaying a “normal” amount of

a certain characteristic and it is these differences that can thus be measured. It is the areas where people exhibit characteristics to a greater or lesser degree than the average person that become apparent to the observer. These characteristics define a person's temperament style, and most people possess two or three components of temperament that are predominant (Pervin, Cervone & John, 2005; Bartram, 2005; Pervin, 2002). These dominant components will have the strongest influence on an individual's needs, goals and behaviour. Being relatively weak on one component can be somewhat compensated for by the presence of another component. In saying this, there is a dynamic balance of these characteristics. Some characteristics may be very weak within a person, whilst other characteristics may be very strong (Mayer, 2005). The implication being that the more a particular behaviour is exhibited, the stronger a characteristic is being represented. However, behaviours can also be exhibited due to a lack of a particular characteristic. Bartram (2004) adds that a person's unique characteristics are regarded as more important than qualifications, training or experience. This is because a person can be trained in order to develop new knowledge and skills. However, a person's attitude, honesty or way of dealing with people are characteristics that are relatively fixed and unchangeable. Therefore it is important to expand the factor of temperament further and consider ways in which to assess or discover an individual's predominant temperament 'style' (Humm & Wadsworth, 1935; Pervin, Cervone & John, 2005).

A component or characteristic of temperament refers to a combination of traits frequently found together and leading to behaviour that is recognisable as characteristic of that particular combination of traits. Further to the definition of personality provided by Pervin, Cervone and John (2005), a trait therefore refers to the consistent patterns of

behaviour, feeling and thought that an individual displays. A trait is a term used to describe a unit of behaviour that cannot be further subdivided, although it can manifest itself differently in different circumstances and in differing combinations with other traits. Behaviour resulting from such combinations of traits can sometimes be referred to as 'typical', however, it is noted that individuals rarely possess all the traits associated with any one characteristic or component and will generally demonstrate (through their behaviour) the possession of traits from several different components. A sub-component therefore refers to a combination of traits found within a component. A trait provides a useful method to summarise how one individual differs from another. Each trait may vary from weak to strong in its manifestations, so that the differences among people are both qualitative as to traits possessed and quantitative as to the strength of each trait. Moreover, the influence of the traits on each other is such that the quantitative differences in the traits themselves are expressed in qualitative differences through observable behaviour (Pervin, Cervone & John, 2005; Humm & Wadsworth, 1935; Humm, 1938).

Barrick and Mount (1991) and Tett, Jackson, and Rothstein (1991) conducted reviews of the relationship between personality and job performance, with particular focus on the Big Five model of personality. Both reviews reported that personality measurement was indeed useful for the prediction of an individual's future on the job performance. Humm and Wadsworth (1943) state that it is critical to understand temperament, because temperament is the combination of emotional tendencies that determine how an individual will react to situations that present throughout life. For example: whether a person is controlled or more emotionally reactive; whether an individual shoulders their responsibilities or evades them; whether they are loyal and trustworthy or unreliable; or

whether they are persistent or easily discouraged. Bartram (2004) adds that assessment is often carried out by an organisation as a method for predicting future on the job performance for both future and current employees, either as a part of a selection and recruitment process or for development and performance management. Performance is often measured by observing the specific behaviours a person displays on an assigned task and then rating them against specific key performance indicators or competencies to identify how appropriate or effective a particular person is at a particular job. Organisations have been using methods such as performance appraisals, reference checking, peer reviews, and other less sophisticated and equally subjective methods to identify an individual's level of performance. However, a proper evaluation of an individual's personal attributes and temperament will also help predict behaviour in any given environment and the validity of personality attributes for predicting job performance is well supported (Bartram, 2004; Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991; Humm & Wadsworth, 1943). Thus, the current investigation utilises a temperament measure that has been used for this purpose.

### **The Humm Wadsworth Temperament Scale**

The Humm Wadsworth Temperament Scale (Humm), developed by Humm and Wadsworth in 1935, was chosen in the present investigation because it has been used for a substantial period of time for the purpose of temperament assessment, particularly with relation to predicting the future on the job performance of an individual. Despite being utilised for over fifty years, there is little established statistical information on the measure, particularly in terms of its application in recent years. Furthermore, the Humm is used in a

commercial environment where factors such as the burden of time required to complete the questionnaire can be a factor. Added to this, the items of the measure have not had a major revision since the measures conception in 1935. Thus given the lack of recent revision information available, the length of the current measure and the time it takes to complete the questionnaire, the Humm was seen as a measure worthy of further investigation.

The Humm is distinct from other measures in that an individual's responses do not generate a specific score or set of scores that fit into a predetermined number of temperament descriptors. In contrast, the Humm does not have a predetermined number of descriptors and as such the Humm has many more possible combinations and various strengths of temperament characteristics when compared with other temperament measures. One of the measure's great assets and point of difference is the ability to measure such a wide range of unique temperament characteristics. The Humm purports to measure seven components and 31 sub-components of temperament. These components and sub-components are rated on a nine point scale (a score of one indicating little or no presence of that component or subcomponent, a score of nine indicating a very strong presence of that component or subcomponent). This gives immense subtlety to the data that the interpreting psychologist has to work with and the possible combinations or styles and the number of variations is too large to contemplate. In this sense, the uniqueness of an individual can be appreciated. The Humm sheds significant light on the strengths and weaknesses (or development needs) of an individual's temperament. This includes, but is not limited to: what motivates an individual; what their stressors and stress reactions may be; how they approach their work; how they are likely to interact with others; how they can best be managed; and how they are likely to manage others (Humm & Wadsworth, 1935; 1941).

A human capital solutions company (the sole proprietor of the Humm), currently utilises the Humm to assess an individual's temperament for the purpose of recruitment and selection, promotion and development, or career guidance, and in the past the Humm has been proved reliable and valid<sup>1</sup>. However, it has been identified that there is some potential for the measure to be revised, improved and updated. Literature on the Humm is quite dated and predominantly ranges from the 1930s to 1950s. The company's founders gained worldwide rights to the instrument in the late 1950s and have successfully used it since as a good predictor of an individual's behaviour.

It is important to have an appreciation of how the current version of the Humm was constructed. The temperament characteristics under consideration first became of interest when observing people with psychological disorders. Around the turn of the century, a European American psychiatrist, Aaron Rosanoff (1927) developed a particularly useful way of looking at human temperament. Rosanoff was interested in clinical or abnormal behaviour, his perspective being that abnormal behaviour is only behaviour exhibited to a degree that impedes proper functioning in a given situation. Rosanoff's belief was that abnormal behaviour is driven by the same core components all humans possess, but to excessive levels and without any control. Rosanoff identified core characteristics and once he could measure them, he had a means by which to understand abnormal behaviour. Many years later an American industrial psychologist, Doncaster Humm identified the theoretical framework as being highly useful for looking at functional people in the workplace with the

---

<sup>1</sup> Investigations into the reliability of the scale found a mean test-retest reliability of .86, an internal consistency of .83, and concurrent validity studies yielded coefficients of .94 and .98 (Kruger, 1938; Humm & Humm, 1944; Smith, Gudmand & Marke, 1958).

view to improve selection and career planning. He identified that there was value in temperament analysis, and in partnership with Guy Wadsworth, a statistician, the Humm Wadsworth Temperament Scale was developed (Humm & Wadsworth, 1935). Humm and Wadsworth applied Rosanoff's theory to a functional 'normal' population, which enabled greater capability in predicting behaviour.

The original standardisation of the Humm consisted of preparing a questionnaire and selecting subjects who displayed known temperament characteristics to whom the questionnaire was administered. Then their responses were analysed to determine the value of each of the items in the questionnaire and the future significance of scores when the measure is administered to unknown subjects at a later date. When the original items were selected for the questionnaire, approximately 2,000 questions were compiled which appeared to have relevance to the traits of the Humm's seven components of temperament: Normal, Hysteroid, Cycloid Manic, Cycloid Depressive, Schizoid Autistic, Schizoid Paranoid and Epileptoid (Humm & Wadsworth, 1935). These components are now known as Normal, Hustler, Mover, Double-Checker, Artist, Politician, and Engineer for obvious commercial and politically sensitive reasons. Once this large number of questions had been compiled, Humm and Wadsworth met with Rosanoff and selected 221 items that seemed most likely to measure the above seven components of temperament. These items were tested on experimental groups and were found to give reasonable results, although not as good as would be necessary if the measure was to be valuable in appraising prospective employees. Subsequently, the items were all reconsidered. Those items which had proved useless in the first trial standardisation were discarded and enough additional items were

included to make a total of 318 questions which constitute the present form of the Humm Wadsworth Temperament Scale<sup>2</sup>.

Humm and Wadsworth (1935) established the original norms from a sample taken from the general community. In later developments, further samples were taken from individuals in employment to develop norms for the industrial community. General population norms were established in 1950 and industrial norms in 1955. The owner of the measure revised the norms for the Australian population in 1966, again in 1977 and in 1999.

Several basic assumptions were necessary when the Humm was created. Firstly, that an individual answering selected questions was, in itself, a sampling of behaviour by which temperamental tendencies could be observed. The reasoning for this being that people of similar tendencies would answer the questions in a similar fashion, while differences in temperamental tendencies would lead to differing responses. The second assumption was that those individuals possessing the temperament characteristics to be measured could be recognised by some other method independent of the questionnaire in order to provide criterion groups for testing the questions, and for the standardisation of the measure. In this case, an alternative method such as behavioural observation could be employed due to the fact that temperament traits can be exhibited quite overtly and predominant components can usually be identified through observation (Humm & Wadsworth, 1935).

---

<sup>2</sup> It is noted that experimental and control groups were used, item analysis was conducted and raw-score norms were established due to the components of temperament not mapping directly on to a normal distribution curve. However, the details of these are beyond the scope of the current study and are therefore not reported on further.

In judging the usefulness of a measure, it is important to know whether or not it is a stable and consistent measure of the variables it is designed to investigate. Previous studies using the Humm have determined the reliability and validity of the current measure (Kruger, 1938; Humm & Humm, 1944; Smith, Gudmand & Marke, 1958). Dysinger (1939) yielded test-retest reliability scores for the seven components significant at the 0.1 % level, with a mean  $r$  of .847, indicating that if a person was to answer the questionnaire a second time, their result would typically be the same as their first results. The effectiveness of the Humm as one of the procedures to be used in personnel appraisals has also been reported favourably by a number of its users. Humm and Humm (1944) confirmed this with research involving one of their clients that yielded a correlation of +0.72 between test results and the ability to predict future performance on the job. A private follow-up study in 1974 replicated these results<sup>3</sup>. In 1997 the owner of the Humm surveyed 56 of their clients regarding the accuracy of the information provided to the client about their employees (based on the employees' results derived from the Humm). There were 225 appraisals in total and the clients surveyed were in a position to observe their employees' behaviour over time. Clients rated the accuracy of the information provided on an employee using a Likert Scale from one to five, one being the information provided was inaccurate with the observed behaviour, and five being the information provided was consistently accurate with the observed behaviour. Of the 56 clients surveyed, 91 percent rated the information provided as consistently accurate with the observed behaviour of their employees. These studies indicate that the Humm was constructed with considerable care to ensure that the measure contains a representative sample of items relating to temperament characteristics

---

<sup>3</sup> For commercial in confidence and privacy reasons further details of these clients and studies cannot be published.

of all types, and that the measure has a definite theoretical basis that can be used effectively by trained psychologists who have a sound understanding of this theory.

Additionally, the Humm provides a gauge of response bias for the total measure and for each of the seven components within the Humm (Humm & Humm, 1944). These 'fake' measures (sometimes referred to as social desirability or impression management), indicate the degree of defensiveness or suggestibility with which the individual has responded to the questionnaire (Dicken, 1963; McCrae & Costa, 1983; Salgado, 1999; and McCrae, 1986). This form of response bias can contaminate the overall test scores. Therefore, among other things, the Humm provides a means for evaluating the extent to which an individual displays this and adjusts their scores accordingly. This phenomenon is supported by Dicken (1963), McCrae and Costa (1983), and McCrae (1986), who suggest that if people respond to the desirability of an item, rather than the content of the item itself, controlling for this response bias should enhance the validity of scores derived from the measure. In the original research by Rosanoff (1927) it was discovered that some of the institutional subjects under-reported their faults to the extent that their responses returned a profile similar to those of normal subjects. Similarly, some of the normal subjects over-reported their faults such that their profiles were similar to those of the institutional subjects. Investigations showed that the former invariably answered predominantly 'No' to the items in the questionnaire, while the latter answered predominantly 'Yes'. In addition, it was found that subjects whose profiles were in agreeance with their case histories tended to distribute their answers fairly evenly between 'Yes' and 'No'. Two measures of response-bias were developed to counter the effect caused by an imbalance between 'Yes' and 'No' responses. Firstly the 'No Count' or number of times an individual responds 'No', and

secondly the profile count or the amount by which the profile positions of all the components except Normal vary from the zero or 'typical' position.

Interestingly, as well as the total scores for each of the components and subcomponents of temperament that are used for the Humm's interpretation, the Humm also provides measures for the accuracy of the information obtained. Firstly, as discussed above, it provides an overall measure for an individual's responses, being the total number of times an individual responds 'No'. Values ranging from 120 to 220 out of a possible 318 responses are deemed an acceptable range for this 'No Count' measure. Secondly, the Humm provides seven individual measures, one for each of the components. These measures are referred to as corrective factors and scores ranging from .75 to 1.75 are deemed as being within an acceptable range. However, less reliable information on one component does not necessarily mean that the whole test for an individual is unusable and this decision is open to the interpretation of a qualified psychologist. When the Humm's measures of response bias are triggered and are deemed significant, meaning that the scores do not fall within the acceptable range for standard interpretation, the psychologist will often administer a second personality measure such as the NEO-PIR or the 16 Personality Factor Questionnaire (16PF) as a confirmatory measure.

The owner of the Humm has been using the measure since the company's beginning, first under licence and then as the proprietary holder when it purchased the rights to the measure on Doncaster Humm's retirement. The measure was brought to Australia from America and re-normed for the Australian general population. The Humm is purported to identify an individual's temperament characteristics and their respective levels

with accuracy, and can make population comparisons to provide useful information that can be used for predicting future performance in the workplace.

There are very few publicly available or published studies examining the Humm from an empirical standpoint, and to present knowledge this is the first study to attempt to reduce the number of items in the questionnaire without decreasing the statistical validity of the measure. Another problem identified by consumers of the Humm is the burden of time it takes an individual to complete the questionnaire due to the number of items being quite large. Therefore one of the goals of the current study was to reduce the number of questions without reducing the statistical validity of the measure. In essence, one of the aims was to make the questionnaire shorter whilst still having an appropriate and acceptable level of statistical validity.

The Humm is a psychological measurement instrument administered as a questionnaire to measure those characteristics which Humm and Wadsworth describe as making up an individual's temperament. Whilst all 318 items of the Humm address issues pertaining to work and life in general, and contribute to the overall response bias measure of the Humm, only 164 of these items load and group into the Humm's seven components and 31 subcomponents. Through an individual's responses to these questions, a profile can be generated which provides information about an individual's temperament across the components and sub-components measured. The responses to the questionnaire were once scored by hand but are now computer scored to produce an output of results that can be interpreted by a trained psychologist. The Humm has consistently maintained its interpreting integrity in that the owners of the measure do not allow other than fully

qualified and accredited psychologists to engage in its interpretation. The owners enforce a number of regulations to ensure this occurs. The owners: only permit the measure to be used for industrial purposes; confine the use of the Humm to psychologists whom the company can readily monitor, namely, psychologists trained by and employed through the owner; require their psychologists to undergo a rigorous six month training programme irrespective of their professional background; and require their psychologists to undergo monthly audits to ensure the ongoing quality of their interpretation skills.

As aforementioned, the possible number of combinations of subcomponents that can occur within the limits of the acceptable range is approximately one billion. One of the measure's great assets is the ability to account for such a wide range of unique temperaments. However, it is important to remember that the results of the Humm, as used by psychologists for the purposes of indicating how a person will perform in a variety of differing employment circumstances, are used in conjunction with other relevant information such as cognitive abilities and previous experience. Through careful interpretation of the seven components and 31 subcomponents by trained and accredited psychologists, a detailed picture of an individual's temperament is generated. The information gathered can shed light on such things as an individual's general potential, motivation, leadership style, business acumen, interpersonal style, work approach, team approach, stress tolerance, level of initiative, and self-confidence. This information is most commonly used for forecasting the future behaviours of an individual and can be utilised for, among other purposes, recruitment, career guidance, team building, career development and consideration for promotion. Currently the Humm is predominantly used as a measure

to identify an individual whose temperament is both suited to the work that is to be completed and the environment within which the individual is to operate.

The Humm's subcomponents were developed by breaking down the seven components of the Humm into a finer and more detailed analysis. This arose from the discovery that any given individual may manifest some of the tendencies associated with a given component, but not necessarily all of the component tendencies. Humm and Wadsworth's first attempt to subdivide the components resulted in 40 subcomponents being created. This was a more detailed breakdown than was justified, since some of the subcomponents were identifiable by too few items. Humm later reduced the number of subcomponents to 31, none of which had less than 12 questions attributed to it.

The current Humm distributes the 31 subcomponents of temperament across the seven components as follows (and illustrated in Figure 1 on p.18): The Normal component has four subcomponents; the Hustler component has six subcomponents; the Mover component has four subcomponents; the Double-Checker component has five subcomponents; the Artist component has five subcomponents; the Politician component has three subcomponents; and the Engineer component has four subcomponents. These subcomponents can be described in more detail, however, for the purposes of the current analysis we will focus on the seven major components only. Whilst the 31 subcomponents give immense subtlety and uniqueness to the data gathered on an individual, knowledge of the seven components of temperament alone can provide an appropriate level of information for management and human resource practices, and it is to these seven components that this report now turns.

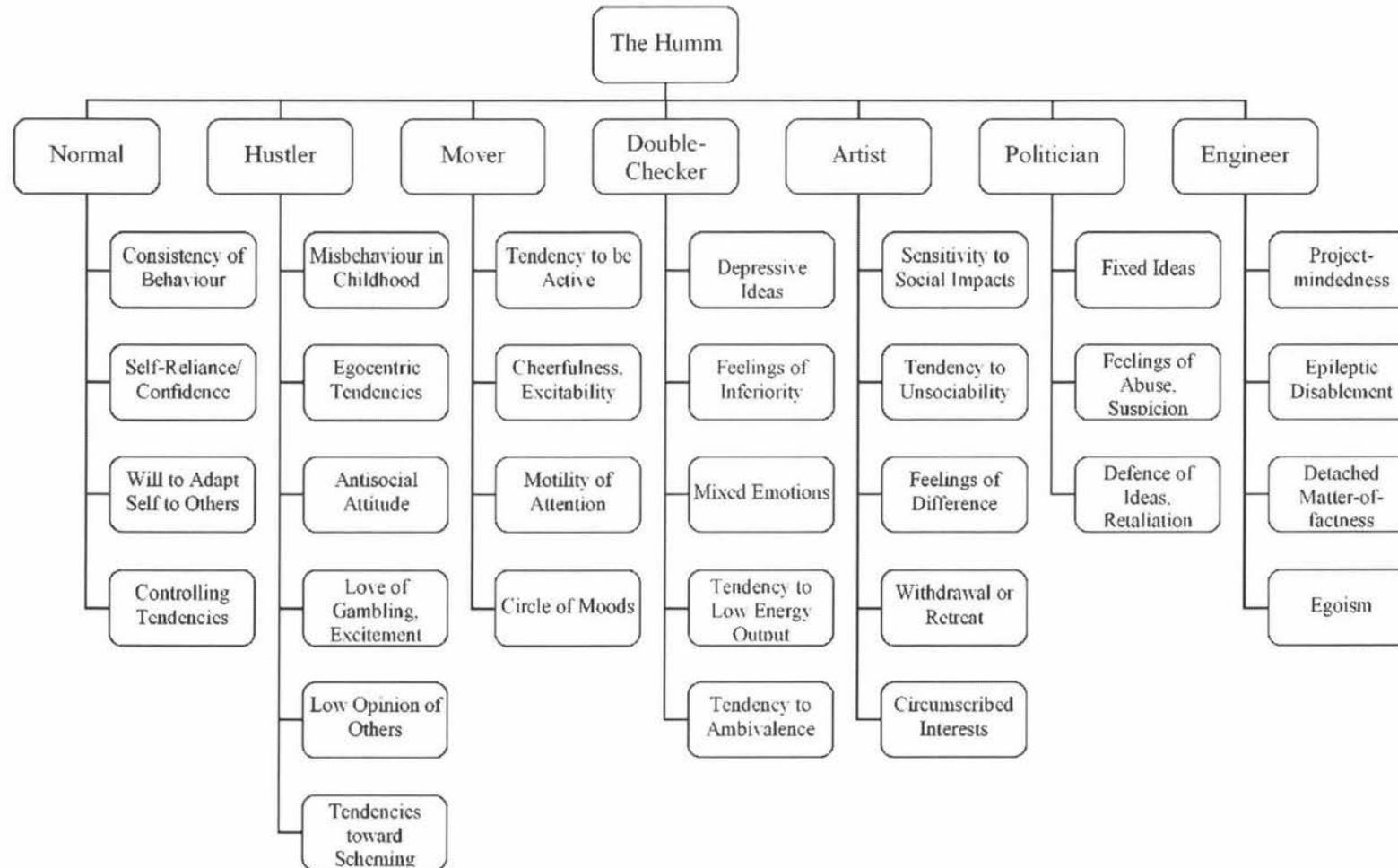


Figure 1: The Seven Components and 31 Sub-components of the Humm

## **The Humm's Seven Components of Temperament**

The descriptions that follow are based on Rosanoff's (1927) work and the later work of Humm and Wadsworth (1935), and refer to the stereotypical behaviour associated with each of the seven (Normal, Hustler, Mover, Double-Checker, Artist, Politician and Engineer) components of temperament within the Humm.

### ***Normal***

The Normal component includes a group of traits or tendencies that provides a certain amount of control over the other six components. It is responsible for the power of self-direction and self-mastery, conservatism and the desire to conform. A person displaying a high level of the Normal component seeks self-improvement and applies a high degree of control over their emotional reactions even in stressful situations. They can also have a conservative attitude to rules and regulations, and can be highly adaptable to social or peer group expectations. This component is frequently described in terms of emphasising its inhibitory and repressive functions and importantly its directive and integrative functions. It not only prevents unfavourable manifestations of the other six components, but also enables the valuable and constructive manifestations of the other six components to present themselves. The Normal component also acts as a measure of self-mastery to evaluate the degree to which an individual's temperament characteristics integrate and is measured by the relationship between the overall score of the Normal component and the scores of the other six components. The Normal component is effective in discriminating between individuals who are masters of themselves and individuals who

give way to impulses and may behave erratically. This measure is so effective in fact, that with a considerable degree of certainty, an individual who has a high level of the Normal component is generally well integrated, while an individual who has a low level of the Normal component is likely to have some difficulty coping, particularly if faced with challenging or difficult situations.

### ***Hustler***

The Hustler component includes the group of traits that leads to a preoccupation with self-interest, the furthering of personal agendas and the satisfaction of personal desires, to the point of not considering the interests and desires of others. The Hustler component includes such attributes as the desire for financial gain, the need for excitement and short-term gratification, the possession of diplomacy, tact and persuasion skills, as well as business acumen and commercial astuteness.

### ***Mover***

The Mover component is responsible for an individual's activity, energy, motility and sociability. Many associated tendencies found in the Mover component include cheerfulness, enthusiasm, and jocularity, responsiveness to others, versatility, hopefulness, and the ability to multi-task.

### *Double-Checker*

Closely related to the Mover component is the Double-Checker component. This component also centres on feelings, emotions and associated manifestations. However the Double-Checker component is responsible for negativity, caution, self-critical behaviour, pessimism, anxiety, empathy, and the manner in which a person makes decisions.

### *Artist*

The Artist component is responsible for shy, sensitive, introspective behaviour. An individual displaying a great deal of the Artist component will be socially sensitive, frequently experiencing some difficulty in expressing their ideas and opinions in face-to-face situations. They are imaginative and creative people, who may be subject to reclusive reactions resulting from feelings of difference. The Artist component includes attributes such as insightfulness, self-consciousness, embarrassment and withdrawal.

### *Politician*

Responsible for ego-defensive behaviour, assertiveness, competitiveness, stubbornness and defensiveness to criticism is the Politician component. Individuals with a high level of the Politician component may be argumentative and can display suspicious, vengeful or aggressive behaviour. They are generally ambitious individuals who are driven by the desire for status, power and prestige.

## *Engineer*

The Engineer component is responsible for systematic, precise, matter-of-fact behaviour, as well as emphasising organisation, procedure, detail and method. Deliberate in approach, people displaying the Engineer component can be quite meticulous, task-orientated, objective, and detail-minded and gain satisfaction through accomplishments.

### **The Humm in relation to other Personality Measurement Instruments**

How does the Humm framework compare with more widely used and published personality measurement instruments? One of the most commonly used sets of traits today is that of the Big Five (Costa & McCrae, 1992). The dimensions of the Big Five are as follows: Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness. Whilst these five factors do not directly align with the seven components of the Humm, some similarities are evident. Neuroticism includes such characteristics as worrying and nervousness and seems most closely aligned to the Double-Checker component of the Humm. The Extraversion trait includes characteristics such as being person-oriented and talkative which seems strongly related to the Mover component. Openness relates to creative and imaginative characteristics, which appear similar to the Artist component. Agreeableness refers to whether a person is cynical, suspicious, vengeful or manipulative and these characteristics seem to align with both the Hustler and the Politician components of the Humm. Finally, Conscientiousness includes characteristics such as being organised, self-disciplined, ambitious and hard working and these appear similar to the Normal and Engineer components of the Humm. The questionnaire that Costa and McCrae (1992)

developed through factor analyses of personality ratings that incorporated the Big Five factors is called the NEO-Personality Inventory Revised (NEO-PI-R). Each of the Big Five factors is further broken down into six facets (as illustrated in Figure 2 on p.24), and eight questionnaire items measure each of these six facets, equating to a total of 240 items in the questionnaire. This is a similar concept to the Humm questionnaire, although subjects indicate for each item the extent to which they agree or disagree, using a five-point rating scale which differs from the forced 'Yes' or 'No' choice for the Humm. Providing further support for the Big Five approach is a similar framework referred to as the Great Eight competency structure as discussed by Bartram (2005). Whilst the relationship between the Great Eight and the Big Five is not exact, the Great Eight does incorporate most of the aspects of the Big Five approach. It appears that frameworks such as the Big Five are similar in concept to that of the Humm. Whilst the components are referred to by different names and the characteristics of the components are grouped slightly differently, the majority of the Humm traits are represented in some way across the overall measure.

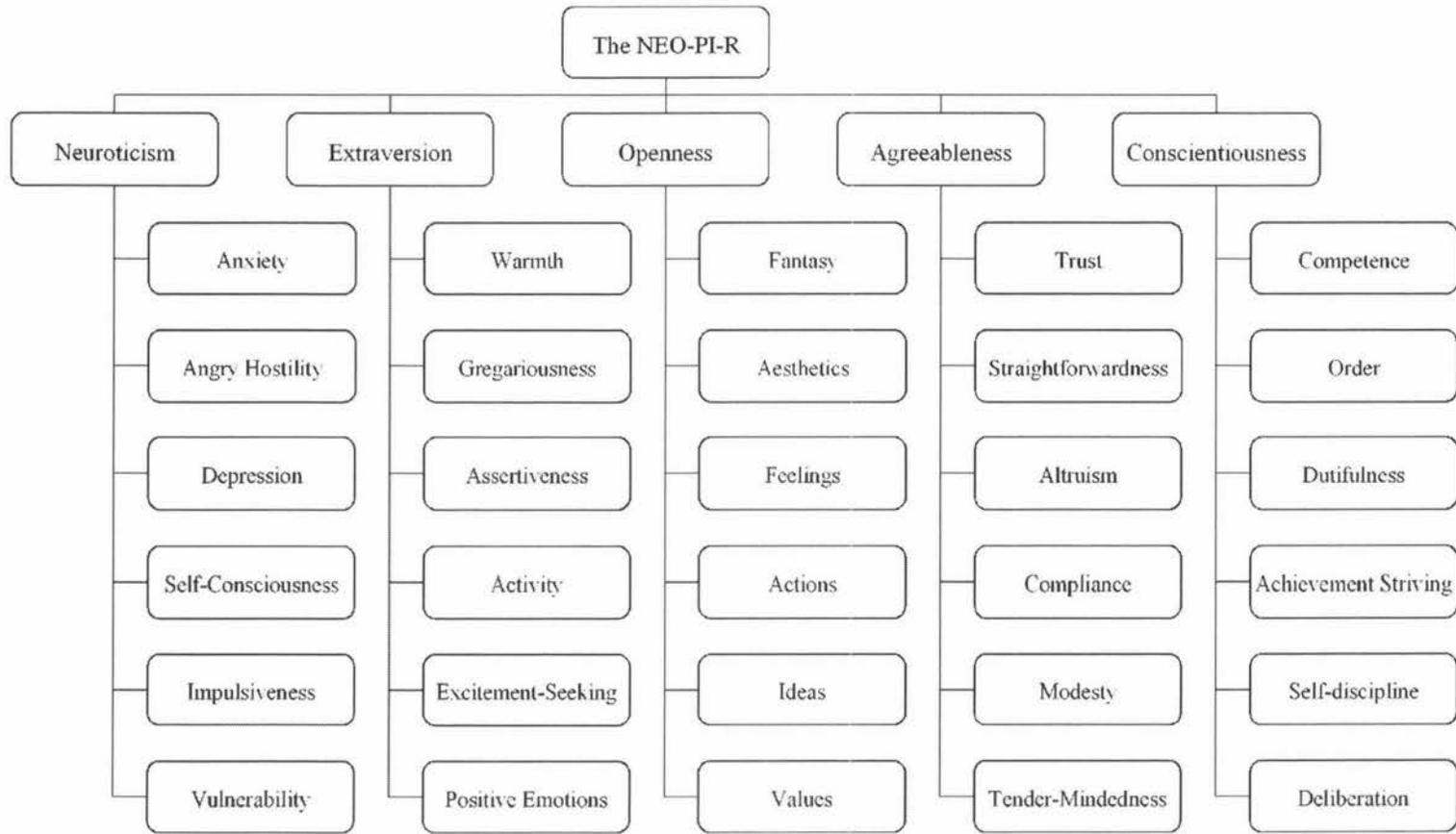


Figure 2: The Big Five Factors and 30 Facets of the NEO-PI-R

## **Evidence for Revision of Psychological Measurement Instruments**

Butcher (2000) discussed guidelines for personality test revision, stating that many psychological measures require updating in order to ensure that their timeliness and effectiveness can be maintained. However, when revising a measure, certain aspects must be maintained to make sure that the revision exercise does not create a new measure altogether. The revised version of a measure must be similar, if not identical to that of the original measure with regards to its structure and configuration. Butcher goes on to say that the revised version of a measure must also be a distinct improvement from the original version, so that the assessment standards of the original version of the measure are raised. Furthermore, Butcher (2000) states that a revision exercise for any measure should be based on and supported by clear empirical justification and rationale, not merely pressure from market forces or other commercial interests. Commercial viability considerations such as the time taken to complete the questionnaire did, in part, drive the current investigation. However, the lack of statistical information available on a measure widely used to assess an individual's temperament, as well as the empirical vulnerability of the Humm, were the key drivers for the present research.

Butcher (2000) suggests that it is also important to gain input from a variety of sources during the revision process, thus qualified Humm experts and users were consulted. To this end, employees from the psychological services team within the owner of the Humm from were asked to contribute to the research, of which six employees responded. In addition, at the conclusion of a revision empirical evidence on the validity of the revised version of the measure is required. Hull, Lehn and Tedile (1991) state that most measures

can indeed have their goodness-of-fit statistics improved through modification post-development and this can sometimes be due to chance fluctuations in the sample data. Therefore, it is important that measures with post-development modifications are always replicated. Thus in the current study a second confirmatory factor analysis was conducted using one half of the data collected to account for this possible effect.

## **Measurement**

It was proposed by Torgerson (1962) that in the social sciences there is a tendency to concentrate on psychological measurement instrument construction, where the means becomes the end of the measurement process. Furthermore, measurement in the discipline of psychology has always been controversial as psychology is not as tangible as other disciplines, such as physics, for example. Some common methods of measurement include: the ordinal assignment of numbers to primary physical qualities, for example, height; counting units which are of the same magnitude; or solving inequalities through ordering and cancellation. The last measurement method is worthy of consideration for applying to the field of psychology (Krantz, Luce, Suppes, & Tversky, 1971; Luce & Tukey, 1964; Torgerson, 1962). Luce (1963) states that fundamental measurement is based on additivity, going on to say that an additive psychological variable has not yet been discovered. Nevertheless, Thurstone (1959) conducted an experiment that showed that additivity of psychological values was indeed possible. However, to adopt classical measurement methods when measuring psychological attributes, additivity must be proved, not simply assumed. Assigning numbers according to rules does not automatically denote that the entity has been measured; rather it only indicates that the entity under investigation has

been classified (Grimm & Yarnold, 1995; Stevens, 1946; Torgerson, 1962). Consequently it is essential to establish the measurement properties of a temperament measure in order to ensure the legitimate and appropriate application of the analyses. Borkenau (2001) adds that there is a distinct difference between measuring a person's abilities, and measuring a person's personality. Measuring abilities involves sampling a person's relevant behaviour, whereas measuring personality is generally based on questionnaires and self-rating scales. Personality measurement most commonly relies on judgmental instruments, that is, the person being assessed makes judgments about their own personality and their typical behaviour (Borkenau, 2001).

To date, the common methods of investigating a psychological construct (including personality and temperament), have taken three forms: The first is that of the total score approach (the summing of subcomponents that have equal weighting); Secondly is the individual score approach (where results for each subcomponent are reported on); and thirdly is the regression approach (which uses multiple regression analysis simultaneously on each of the subcomponents) (Hull, Lehn & Tedile, 1991).

There are advantages and disadvantages for each of the above three methods. The total score approach may also include subcomponents that are weak or useless, which can lower the overall effectiveness of a construct. Whilst the individual score approach overcomes this problem, it also introduces ambiguity and complexity due to the analyses of multiple subcomponents. The regression approach also identifies the unique effects of each of the subcomponents. However, this approach can be limited by multicollinearity (where strong relationships exist between subcomponents) and this may cause estimated regression

coefficients to become unstable. The regression approach can also suffer from differential reliability problems whereby less important subcomponents may be measured as being more reliable than more important subcomponents that have been poorly measured (Hull, Lehn & Tedile, 1991).

### **Thurstone's Method of Paired Comparisons**

A promising method for applying psychophysical measurement theory to psychology rests with Thurstone's method of paired comparisons (Coombs, Dawes & Tversky, 1970; Mosteller, 1963; Thurstone, 1927b, 1927c, 1931; Torgerson, 1962). Thurstone's original experiment, based on the seriousness of different crimes, provided a list of paired offences to university students and asked them to rate which was the most serious offence of each pair. The responses enabled the construction of a frequency matrix containing the relative frequency of the preferred choice in each pair.

The method of paired comparisons assumes that a stimulus arouses a discrimination process in an individual that creates a value that can then be placed along a psychological continuum (Edwards, 1957; Michell, 1990; Thurstone, 1927a). Different individuals vary in their opinion, and therefore their point of discrimination on a particular stimulus along a psychological continuum. However, the responses of a number of individuals will converge on a normal distribution as the number of participants increases, as predicted by the central limit theorem (Howell, 1997).

Thurstone (1927a) asserted that a single question alone cannot provide sufficient information for the construction of a psychological measurement instrument. In contrast, a succession of paired questions measuring the psychological distance between pairs of stimuli can provide data that can be mapped on a psychological continuum, thus enabling psychological values to be calculated for each stimulus. The proportion of respondents preferring one or the other paired stimuli from the frequency matrix is then used to create a proportion matrix. The proportion matrix is converted to a standardised matrix, providing a consensus set of psychological values for the stimuli under consideration.

While Thurstone's method is straightforward and succinct, unidimensionality and quantitativity of the psychological attribute is assumed, but not established (Michell, 1990). Luce (1963) argues that Thurstone's equal variance assumption has not yet been proved. However, Lord and Novick (1968) suggest that Thurstone's method can lead to an interval scale, which is enough for the current purposes. Although Thurstone's assumption of equal standard deviations of stimuli is sometimes challenged, Moesteller (1963) argues that Thurstone's assumption is reasonable and any arguments to the contrary are not enough to discourage further use of this method.

In terms of the present investigation, by conducting a thorough conceptual analysis of the items in the Humm, the aim of developing a better selection of items for each of the seven temperament components can then be addressed. The application of Thurstone's methods providing psychological values for items can then be applied to potentially achieve increased reliability of the revised version of the measure (Thurstone, 1927b). The application of Thurstone's psychometric values to the evaluation of items in the Humm can

provide a quantitative foundation for the revision of the measure. Establishing the quantitative attributes of the stimuli plausibly places the Humm construct on a more sound theoretical foundation than is presently the case. Thurstone's method of paired comparisons appears to be a promising tool for evaluating some of the psychological attributes underlying the Humm. The current study aims to utilise this method to provide psychological values that can be applied to ascertain a more valid set of items, and therefore create an empirically stronger psychological measurement instrument.

### **Aims and Rationale of the Current Study**

In sum, the present study addresses three main issues, which to present knowledge no previously published study has attempted to investigate. The first is the evaluation of the psychometric properties of the Humm Wadsworth Temperament Scale (Humm), with the view of reducing the number of items in the measure without reducing the empirical validity of the measure. The second is the application of quantitative theory, in this case Thurstone's method of paired comparisons, to evaluate the items in the Humm conceptually and to possibly improve construct validity by choosing appropriate items as required by the psychological domain under investigation (Anastasi & Urbina, 1997). The final issue is to utilise confirmatory factor analysis to refine the items identified conceptually in order to create the most statistically appropriate set of items possible in the present circumstances. The above aspects of this study, based on proven psychometric methods, can then be integrated into future revisions of the Humm (and similar temperament measures), and provide a solid foundation for future enhancements.