

CROSS-NATIONAL CONSTRUCT EQUIVALENCE OF SCHOOL-AGE
CHILDREN'S TEMPERAMENT TYPES AS MEASURED BY THE STUDENT
STYLES QUESTIONNAIRE

By

NICHOLAS F. BENSON

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2005

Copyright 2005

by

Nicholas F. Benson

ACKNOWLEDGMENTS

I would like to acknowledge the contributions of my committee members. First, this project would not have been possible without my dissertation chair, Dr. Thomas Oakland. His interests in temperament and international outlook on psychology were instrumental to the formulation of this project. His competencies as scholar and mentor were instrumental to the completion of this project. I consider myself exceedingly fortunate to have had the benefit of his guidance and expertise. Second, Dr. Mark Shermis provided invaluable methodological guidance that enabled me to address the proposed research questions. I am indebted to him for the methodological and statistical knowledge I acquired under his tutelage. Third, I would like to thank Drs. Jennifer Asmus and Phillip Clark for their commitment to serve on my committee and their contributions to this project. I also would like to acknowledge my family and the importance of their love. In particular, I would like to thank my wife Grace for affectionately sharing advice and offering encouragement. At last, I would be remiss not to thank my children for the joy they bring to my life.

TABLE OF CONTENTS

| | <u>page</u> |
|--|-------------|
| ACKNOWLEDGMENTS | iii |
| LIST OF TABLES | vi |
| LIST OF FIGURES | ix |
| ABSTRACT..... | x |
| CHAPTER | |
| 1 INTRODUCTION | 1 |
| 2 REVIEW OF THE LITERATURE | 4 |
| Description and Delimitation of Temperament | 4 |
| Ancient Perspectives on Temperament | 7 |
| Hippocrates..... | 7 |
| Galen..... | 8 |
| Immanuel Kant | 8 |
| Eastern Scholars | 9 |
| Modern Perspectives on Temperament | 10 |
| Wundt's Typology..... | 10 |
| Jungian Typology..... | 10 |
| Personal Disposition Theory | 12 |
| Trait Theorists | 13 |
| Goodness-of-Fit Theory | 18 |
| Biological Approaches | 22 |
| Myers/Briggs Type Theory | 27 |
| Keirsian Theory | 28 |
| SSQ Theory | 30 |
| Trait Theory and Type Theory Relationships | 40 |
| Practical Uses of Temperament Concepts | 42 |
| Education..... | 44 |
| Vocations..... | 47 |
| Considerations in the Generalizability of Temperamental Constructs | 48 |
| Developmental Considerations..... | 48 |
| Gender Considerations | 50 |
| Cultural Considerations | 50 |

| | |
|---|-----|
| Measurement..... | 57 |
| Psychological Tests | 57 |
| Validity | 58 |
| Test Adaptation | 59 |
| Purpose of the Study..... | 66 |
| Research Questions..... | 67 |
| | |
| 3 METHODS | 68 |
| | |
| Participants | 68 |
| Instrumentation | 69 |
| Procedure | 71 |
| Data Analysis Procedures | 72 |
| Factor Analysis..... | 74 |
| Multidimensional Scaling and Cluster Analysis | 75 |
| Covariance Structure Analysis Techniques..... | 76 |
| Data Analysis Procedures used for the Present Study | 77 |
| | |
| 4 RESULTS | 81 |
| | |
| 5 DISCUSSION..... | 97 |
| | |
| Implications | 99 |
| Cross-National Implications..... | 99 |
| Methodological Implications..... | 99 |
| Test Validity Implications | 101 |
| Limitations..... | 103 |
| Directions for Future Research..... | 103 |
| | |
| APPENDIX | |
| | |
| A LISTING OF PARCEL ITEMS | 105 |
| | |
| B STANDARDIZED REGRESSION WEIGHTS AND LATENT FACTOR CORRELATIONS FOR THE PARCEL MODEL..... | 106 |
| | |
| C CONTENT OF MODIFIED PARCELS | 112 |
| | |
| D STANDARDIZED REGRESSION WEIGHTS AND LATENT FACTOR CORRELATIONS FOR THE MODIFIED MODEL..... | 115 |
| | |
| LIST OF REFERENCES..... | 121 |
| | |
| BIOGRAPHICAL SKETCH | 132 |

LIST OF TABLES

| <u>Table</u> | <u>page</u> |
|---|-------------|
| 2-1 Temperament categories proposed by Thomas and Chess..... | 21 |
| 2-2 Temperament constellations proposed by Thomas and Chess..... | 22 |
| 2-3 Important conceptual differences between trait and type theories..... | 41 |
| 4-1 Comparison of fit for the parcel model. | 90 |
| 4-2 Fit of the parcel model, multi-country analysis..... | 91 |
| 4-3 Comparison of fit for the modified model. | 93 |
| 4-4 Fit of the modified model, multi- country analysis..... | 94 |
| 4-5 Fit for identified three-factor models. | 95 |
| 4-6 Fit for identified hierarchical models..... | 96 |
| 4-7 Comparison of competing models..... | 96 |
| A-1 Parcel items for EI (extroverted-introverted)..... | 105 |
| A-2 Parcel items for PM (practical-imaginative)..... | 105 |
| A-3 Parcel items for TF (thinking-feeling). | 105 |
| A-4 Parcel items for OL (organized-flexible)..... | 105 |
| B-1 Standardized regression weights for the Australian sample..... | 106 |
| B-2 Standardized regression weights for the Chinese sample. | 107 |
| B-3 Standardized regression weights for the Costa Rican sample..... | 107 |
| B-4 Standardized regression weights for the Gazan sample..... | 108 |
| B-5 Standardized regression weights for the Nigerian sample. | 108 |
| B-6 Standardized regression weights for the Philippine sample..... | 109 |

| | | |
|------|---|-----|
| B-7 | Standardized regression weights for the USA sample. | 109 |
| B-8 | Standardized regression weights for the Zimbabwe sample. | 110 |
| B-9 | Correlations among the four bipolar styles for the Australian sample..... | 110 |
| B-10 | Correlations among the four bipolar styles for the Chinese sample..... | 110 |
| B-11 | Correlations among the four bipolar styles for the Costa Rican sample. | 110 |
| B-12 | Correlations among the four bipolar styles for the Gaza sample. | 111 |
| B-13 | Correlations among the four bipolar styles for the Nigerian sample. | 111 |
| B-14 | Correlations among the four bipolar styles for the Philippine sample. | 111 |
| B-15 | Correlations among the four bipolar styles for the USA sample. | 111 |
| B-16 | Correlations among the four bipolar styles for the Zimbabwe sample. | 111 |
| D-1 | Standardized regression weights for the Australian sample..... | 115 |
| D-2 | Standardized regression weights for the Chinese sample. | 115 |
| D-3 | Standardized regression weights for the Costa Rican sample..... | 116 |
| D-4 | Standardized regression weights for the Gazan sample. | 116 |
| D-5 | Standardized regression weights for the Nigerian sample. | 117 |
| D-6 | Standardized regression weights for the Philippine sample..... | 117 |
| D-7 | Standardized regression weights for the USA sample. | 118 |
| D-8 | Standardized regression weights for the Zimbabwe sample. | 118 |
| D-9 | Correlations among the four bipolar styles for the Australian sample..... | 118 |
| D-10 | Correlations among the four bipolar styles for the Chinese sample..... | 119 |
| D-11 | Correlations among the four bipolar styles for the Costa Rican sample. | 119 |
| D-12 | Correlations among the four bipolar styles for the Gaza sample. | 119 |
| D-13 | Correlations among the four bipolar styles for the Nigerian sample. | 119 |
| D-14 | Correlations among the four bipolar styles for the Philippine sample..... | 119 |

D-15 Correlations among the four bipolar styles for the USA sample.120
D-16 Correlations among the four bipolar styles for the Zimbabwe sample.120

LIST OF FIGURES

| <u>Figure</u> | <u>page</u> |
|---|-------------|
| 2-1 Keirsian types and their derivation..... | 30 |
| 4-1 Sixty-nine item Student Style's Questionnaire model. | 82 |
| 4-2 Parcel model..... | 83 |
| 4-3 Modified model..... | 84 |
| 4-4 Three-factor model..... | 86 |
| 4-5 Hierarchical model..... | 87 |
| 4-6 Rationale for analyses. | 89 |

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

CROSS-NATIONAL CONSTRUCT EQUIVALENCE OF SCHOOL-AGE
CHILDREN'S TEMPERAMENT TYPES AS MEASURED BY THE STUDENT
STYLES QUESTIONNAIRE

By

Nicholas Benson

May 2005

Chairman: Thomas Oakland
Major Department: Educational Psychology

Interest in developing tests that can be used internationally has flourished in the last few years. When developing tests for this purpose, it is important to establish cross-national equivalence for the psychological constructs of interest. Measurement of temperament is an important endeavor with international appeal. Information related to temperament can be used to facilitate growth, guide important life decisions, improve individual performance, promote the prevention and intervention of problem behavior, and foster social relationships. Our study first reviewed historical perspectives on temperament constructs. We then explored recent perspectives including type theory, trait theory, and biological perspectives. Practical uses of temperament concepts, considerations in the generalizability of temperament constructs, and measurement issues also were reviewed. Our study was designed to examine the cross-national construct equivalence of school-aged children's temperament types as measured by the Student Style's Questionnaire (SSQ). The SSQ is a

forced-choice self-report measure of children's temperament. Development of the SSQ was driven by Jung's theory of temperament as augmented by Myers and Briggs. Results of our study support the conclusion that the SSQ measures the same four bipolar dimensions or styles among children from eight countries: Australia, China, Costa Rica, Gaza, Nigeria, the Philippines, The United States, and Zimbabwe.

CHAPTER 1 INTRODUCTION

People have attempted to describe and explain individual differences in behavior throughout recorded history. The notion of individual differences is not necessarily judgmental or pejorative. Humans are social creatures. Varied social roles are needed to allow social groups to survive. No one behavioral style is best suited for all social roles.

Ironically, the notion of individual differences also implies the existence of similarities. As noted by Bouchard (1996; p.24), an internationally recognized expert in human behavior genetics, "The fact is that in spite of varying cultural content and dramatic differences in behavioral expression, human beings are remarkably similar with respect to their evolved psychological mechanisms." Bouchard suggests that these psychological mechanisms were born of and refined by the process of natural selection. In other words, specific psychological mechanisms produce behavioral styles that have yielded more frequent mating and greater rates of survival, and consequently these mechanisms have been perpetuated. Although a psychological mechanism may yield behavioral styles that much of society views as undesirable, this mechanism may characterize a significant percentage of the population because the mechanism has had survival value (e.g., has contributed to society).

Throughout history, scholars have developed behavioral constructs (such as temperament) based on behavioral similarities. Constructs allow people to organize and classify, so that they can better make sense of the world. Although temperament concepts

are as old as recorded history, temperament is a relatively new area of scientific research. Temperament has been subject to scientific study for a little more than 50 years.

While researchers strive to develop sound theories of temperament, the need for suitable measures of these constructs persists in research and applied settings (e.g., school, vocational, and therapeutic settings). Test users want the best information available, so that they can describe past and current behaviors, attempt to explain why certain behaviors occur, and try to manage behaviors so they are beneficial to the individual and to the community (Thayer, 1995). Information related to temperament also can be used to facilitate growth, guide important life decisions, improve individual performance, promote the prevention and intervention of problem behavior, and foster social relationships.

The demand for tests to be used cross-nationally has increased in recent years. Temperament tests are no exception, given the usefulness of information obtained from these tests. However, importing and adapting tests to make clinical, educational, or other important life decisions without first examining the comparative validity of test interpretations across countries is inappropriate.

Recently, as evidenced by events such as the 1999 International Conference on Test Adaptation, a number of scholars have actively engaged in studying and refining the process of adapting tests for use cross-culturally and cross-nationally. However, studies using factor analysis or confirmatory factor analysis to investigate the cross-national construct equivalence of temperament measures have yet to be published. Moreover, no evidence was found of attempts to establish the cross-national construct equivalence of either temperament types or personality traits for children. This dearth of research is

partly because of the complex nature of temperament and personality and the tests that assess them. These tests may be objective or projective; may measure unipolar or bipolar variables; and may require responses that are choices rather than right or wrong answers. These complexities hinder the process of acquiring data and comparing test validity across groups (Stafford, 1994). Traditional bias definitions (e.g., slope and intercept bias) and methodology used to establish the boundaries of test interpretations are largely inapplicable to the study of temperament and personality (Stafford, 1994).

The purpose of our study was to examine the construct equivalence of a measure of children's temperaments cross-nationally. Our study examined whether the Student Styles Questionnaire (SSQ), as used in 8 countries, has etic (cultural-universal) qualities. Thus, our study explored whether the SSQ measures the same temperament qualities when used with children in various cultures. If evidence suggests that the SSQ measures the same temperament qualities across countries, then children's temperaments, as measured by the SSQ, may be attributes that transcend cultures. If the evidence suggests otherwise, SSQ interpretations for children outside the United States must consider differences in the constructs underlying temperament preferences.

Our study addressed the following three questions:

- **Specific aim 1.** Does the SSQ measure the same four bipolar dimensions among children from Australia, China, Costa Rica, Gaza¹, Nigeria, the Philippines, the United States, and Zimbabwe?
- **Specific aim 2.** If not, what is the structure of the SSQ in these countries?
- **Specific aim 3.** Are the intercorrelations of the dimensions similar across countries?

¹ Although Gaza may not be considered a country at this time, for convenience it will be referred to as such throughout this document.

CHAPTER 2 REVIEW OF THE LITERATURE

Student styles theory guided the development of the Student Styles Questionnaire (SSQ). Data from the SSQ were used in our study. Temperament and issues that pertain to its measurement are discussed. The purpose of the study is elaborated and research questions addressed are set forth.

Description and Delimitation of Temperament

Temperament concepts pertain to consistencies that can be observed in people's attitudes, preferences, affect, and styles of behaving. Temperament is thought to influence how people choose the path their lives will take and how they react to those experiences that life sends their way. However, temperament generally is not viewed as having a deterministic influence. Instead, the concepts assume a bi-directional interplay between an individual's constitutional traits and environment.

Various temperament theories exist. When adopting a theory of temperament, one makes a number of assumptions about human behavior. Theories and their inherent assumptions can constrain research perspectives. Therefore, views of temperament differ according to the purpose of the researchers (e.g., clinical applications versus finding the biological underpinnings of behavior). Researchers adopt a theory that allows them to generate research and test their hypotheses. Thus, each theory needs a unique definition. Consequently, there is no consensus definition of temperament.

The term temperament can be thought of as a broad rubric that subsumes specific concepts (Goldsmith & Rieser-Danner, 1986). Despite the lack of a consensus definition,

there are convergent ideas regarding temperament (Goldsmith, Buss, Plomin, Rothbart, Thomas, Chess, Hinde, & McCall, 1987). Some of the least contentious ideas held by temperament scholars are as follows:

- The term temperament refers to a group of related traits rather than a single trait.
- Temperament refers to the issue of individual differences rather than species-general characteristics.
- Temperament traits have biological underpinnings.
- The concept emphasizes continuity in behavioral tendencies.
- Despite the emphasis on continuity, temperament commonly is viewed as malleable.
- The concept does not refer to discrete behavioral acts. Rather, temperament traits refer to biases for certain moods, attitudes, and dispositions that affect the probability that behaviors will occur.

Although the above points of agreement and others can be delineated, points of divergence may be more striking. Although temperament research is rapidly evolving (Rothbart & Bates, 1998; Teglasi, 1998), competing theories of temperament persist. Consequently, temperament researchers continue to hold divergent assumptions and beliefs about their subject matter.

Although researchers hold a plethora of divergent ideas, they all appear to stem from two major points of divergence. The first major point of divergence pertains to the boundaries of temperament (Goldsmith et al., 1987). Researchers disagree as to which behaviors constitute temperament. Criteria proposed include the style of the behavior, the relationship of the behavior to emotional systems, the stability of the behavior, and the heritability of the behavior (Goldsmith et al., 1987). The second major point of divergence pertains to the number of temperament dimensions or traits (Goldsmith et al., 1987). This issue typically is addressed empirically using some form of mathematical

factor extraction and/or model-fitting technique. Although researchers have addressed this issue, the unclear boundaries of temperament cause disagreement on this point.

Researchers tend to rename temperament variables, despite their similarity (Rothbart, 1999). Researchers may rename variables in order to highlight the ideas that make their research unique and important, rather than focus on recurring themes in the literature. For example, Kagan (1989) and Thomas and Chess (1977) both hypothesized temperament variables based on response to novelty. Kagan named his variables behavioral approach and behavioral inhibition, while Thomas and Chess named approach or withdrawal variables based on the initial response to new stimuli, and an adaptability variable based on the ease of responses to new or altered situations. Renaming of variables also may result from the interdisciplinary nature of temperament research. Scholars from diverse disciplines such as behavioral genetics, education, psychophysiology, developmental psychology, personality theory, psychosomatic medicine, and clinical psychiatry (Campos, Barrett, Lamb, Goldsmith, & Stenberg, 1983; Goldsmith et al., 1987) contribute to the temperament literature. Researchers may be unaware of the contributions of other disciplines. Whatever the reason, renaming variables leads to confusion and discontinuity in research.

Individual temperament traits are embedded in the context of other temperament traits or variables that constrain their influence (Teglasi, 1998). Moreover, temperament is embedded in the context of personality variables that may mediate or moderate the influence of temperament variables.

Personality generally is described as a broader construct than temperament. For example, personality has been conceptualized in the following ways:

- As a combination of temperament traits plus ability traits or intelligence (Eysenck & Eysenck, 1985)
- As a combination of temperament and dynamic motivational traits that arise in response to the environment (Cattell & Kline, 1977)
- As a mutual interaction of environmental events, behavior, and personal factors (Bandura, 1977)
- As a subset of personality (Buss & Plomin, 1984; Strelau, 1987)
- As the foundation of personality (Kluckhohn & Murray, 1953)

As with temperament, there is no consensus definition of personality. Because temperament studies and personality studies often examine the same questions, distinctions between the two constructs may be arbitrary. It is difficult to define relationships between two constructs that have dynamic, evolving definitions.

Ancient Perspectives on Temperament

Pre-scientific temperament theories (although based on intuition and casual observations) have endured for thousands of years and continue to exert a powerful influence on temperament research, applied practice, and society in general. The most influential Western perspectives have been those offered by Hippocrates, Galen, and Kant. These perspectives and influential perspectives of Eastern scholars are discussed next.

Hippocrates

The first prominent theory of temperament emerged from ancient Greek culture. In the 4th century B.C., Hippocrates, regarded as the father of medicine, proposed a four humor theory that became the early forerunner of modern neurochemical theories. These humors (i.e., yellow and black bile, blood, and phlegm) were thought to create opposition between warm and cool and between dry and moist bipolar bodily qualities. Hippocrates

attributed differences in rationality, emotionality, and behavior to the balance of these humors.

Galen

Galen of Pergamon, a second-century physician, elaborated on Hippocrates' theory (Kagan, 1994). Galen proposed nine temperament (a word derived from the Latin "to mix") types, derived from Hippocrates's four humors. The ideal personality type was believed to result when the four humors combined to form an exquisite balance of the following bodily qualities: warm, cool, dry, and moist. Four less-ideal temperament types resulted from the dominance of one of the four bodily qualities. The final four types that were described as temperamental categories resulted from one pair of qualities dominating the complementary pair (e.g., cool and dry dominating warm and moist). The four temperament categories are as follows: melancholic, sanguine, choleric, and phlegmatic. The melancholic type (cool and dry) has too much black bile, which makes one sad and anxious. The sanguine type (warm and moist) has primarily blood, which makes one enthusiastic and pleasant. The choleric type has too much yellow bile, which makes one irritable and angry. The phlegmatic type has too much phlegm, which makes one's emotions and actions slow.

Immanuel Kant

Hundreds of years after Galen's work, Immanuel Kant, an influential Prussian scientist and philosopher, included a chapter on temperament in his 1798 book *Anthropologie*. Kant believed that psychology could not be an experimental science. He proposed that anthropology could serve as an alternative way of studying how people actually behave and could provide the information necessary to predict and control

human behavior. Kant's book was widely read, and his ideas on temperament were influential in Europe (Eysenck & Eysenck, 1985).

Kant, like Galen, believed that humors formed the basis for temperament. However, Kant believed that blood was the most significant component. He proposed four independent temperaments: sanguine, melancholic, choleric, and phlegmatic. The sanguine person is carefree and sociable. The melancholic person is anxious, and eludes happiness. The choleric person is easily annoyed, and suffers most when others refuse to follow his orders. The phlegmatic person is reasonable, persistent, and acts on principle rather than instinct. To Kant, temperament refers to the energetic (choleric vs. phlegmatic) and the emotional (sanguine vs. melancholic) characteristics of behavior.

People can be assigned to one of the four independent types based on their dominant behavioral characteristics. According to Kant, there are no compound temperaments (e.g., melancholic-phlegmatic).

Eastern Scholars

Ancient Eastern scholars, unlike Western scholars, apparently did not use stable constructs such as temperament when developing explanatory theories for behavior (Kagan, 1994). For example, the Chinese theory of human nature (held by the Chinese for at least 2 millennia before Galen expanded on Hippocrates's theory) held that universal energy (ch'i) exists. Ch'i is regulated by the complementary relationship between the forces of yin (passive and completing) and yang (active and initiating). The balance of yin and yang was thought to regulate physiological and psychological functioning. The energy of ch'i always is changing, and therefore does not describe stable properties.

Ancient Hindu Indians also proposed a theory of human nature that ignored stable, constitutional factors. The Hindus believed that environmental factors (specifically air,

water, and geographic location) influenced bodily substances (specifically spirit, phlegm, and bile); and that this interaction was responsible for observed behavioral consistencies (Garrison & Earls, 1987).

Modern Perspectives on Temperament

Wundt's Typology

The founder of experimental psychology, Wilhelm Wundt, proposed the modern view of temperament in 1903 (Eysenck & Eysenck, 1985). Like Galen and Kant, Wundt's typology offered four temperaments. However, Wundt changed the uni-dimensional, categorical system offered by Galen and Kant to a quantitative, two-dimensional system. In Wundt's system, people can occupy any position (and, moreover, any combination of positions) on the energetic and emotional dimensions offered by Kant. Wundt labeled the dimensions strong emotions as opposed to weak emotions, and changeable as opposed to unchangeable. The position that people occupy in Wundt's two-dimensional system is determined by the strength and temporal characteristics of their emotions.

Jungian Typology

Carl Jung, a psychiatrist and psychoanalyst, developed an interest in temperament while exploring his theoretical beliefs and how they differed from those of other influential psychoanalysts such as Freud and Adler. Jung found his interest in the delimitation and organization of psychological processes to be a prominent difference. Jung detailed these psychological processes in *Psychological Types* (Jung, 1971).

According to Jung, differences among people can be conceptualized into two types: individual and group. Conceptually, Jung's psychological types are group differences.

Groups of people with a particular psychological type hold similar attitudes toward the world that determine and limit their judgment.

Jung believed that Galen's classifications were based on affect. Galen seemingly based his classifications on affect because affect is the most common and striking feature of behavior. To Jung, affect constrains attitudes of the conscious mind, and consequently constrains free choice. Thus, Jung was interested in unconscious mental processes such as attitudes and values.

Jung believed that one should strive to realize one's self. Jung's concept of self refers to a broad array of psychic phenomena that comprise people's inner world. In realizing the self, people transcend all opposing forces, so that every aspect of their personality is expressed equally. He believed that opposing forces produce energy. With no opposition, there is no energy. His theory focuses on those opposing forces that he believed have the most empirically robust impact on the function of people's attitudes.

Jung believed that the principle of opposites governs attitudes. Opposing forces can be thought of as bi-polar dimensions, as everyone has both opposing forces. The dimensions delineated by Jung arise from three qualities:

- Where one derives energy
- A focus on practical details versus theory
- Reliance on thoughts versus feelings when making decisions

The first quality results from the tendency of the libido (by which Jung means more than sexual energy) to be directed to the inner self (introversion) or to objects of interest (extraversion). People develop a primary mechanism attitude-type as they develop (i.e., either introversion or extraversion) based on this quality. People develop secondary mechanisms or function-types based on the second and third qualities. Introversion and

extraversion are expressed via four function-types: thinking, feeling, sensing, and intuiting. Thinking and feeling are considered rational functions, because they use reason, judgment, abstraction, and generalization. Sensing and intuiting are considered irrational functions, because they are based on intensity of one's perception. Jung's psychological types are formed by combining an individual's attitude-type with his or her function-types.

According to Jung, people have a tendency to strive for psychic equilibrium. They delude their judgment so they can compensate for the psychic imbalance created by the predominance of their psychological type. The inclusion of this tendency complicates Jung's theory and makes the interpretation of types difficult, which is why no modern psychologist has adopted his theory in its entirety (Eseyneck & Eseyneck, 1985).

Jung was aware that, for the most part, his typology was not understood by others. Jung also was aware that critics tended to assume that his types were not supported by empirical data. Moreover, in the forward to the Argentine edition of his book, he noted that, even in medical circles, an opinion exists that his method of treatment consists of fitting patients into his system and giving them corresponding advice. In the 1937 forward to the seventh Swiss edition, Jung countered that his typology resulted from years of experience as a practicing psychotherapist (experience that is not available to academic psychologists). Thus, Jung believed that sufficient experience as a psychotherapist was needed to fully understand the meaning and value of his typology.

Personal Disposition Theory

Like Jung, Gordon Allport also was greatly influenced by Freud. Allport, who was born in Indiana, managed to arrange a meeting with Freud in Vienna. Allport was immediately disenchanted by what he perceived as Freud's over-analysis of their

conversation. He believed that the determinism of human behavior is much more complex and varied than the explanations offered by Freud's theory. Allport's writings, especially his 1937 book *Personality, a Psychological Interpretation*, did much to bring an awareness of and interest in personality and temperament to the United States (Lombardo & Foschi, 2002).

Allport focused on the study of personal traits. He later changed the term personal traits to personal dispositions to emphasize the importance of unique, individual characteristics. That is, he believed that knowing other people's traits does not necessarily allow you to understand a particular person. Nevertheless, he did believe that some traits are common to a culture, and are recognized and named by people in the culture (e.g., extraverts and introverts). Common traits are reliably observed and normally distributed. According to Allport, people possess a number of traits. Various traits differently influence behavior. Traits that are sufficiently pervasive to affect all behavior he called cardinal dispositions. He also believed that people typically have five to ten highly characteristic tendencies, which he referred to as central dispositions. He also believed people have specific, focused tendencies that occur only during specific situations. He referred to these traits as secondary dispositions. Psychologists may now refer to them as states.

Trait Theorists

Contrary to Allport's theory and his belief in the individuality of people, researchers such as Raymond Cattell (Cattell & Kline, 1977), Hans Eysenck (Eysenck & Eysenck 1985), and J.P. Guilford (Guilford, Zimmerman, & Guilford, 1976) believed that specification equations can be devised and people's behavior predicted using temperament traits. These researchers shared an interest in using mathematical methods

to study personality. These contemporaries are classified as state-trait theorists. To them, personality has two major aspects: temperament and intelligence. We focused only on the temperament aspect of their work. Armed with their knowledge of mathematics and methodology, trait theorists have done much to improve the scientific rigor of temperament investigation.

Trait theorists shared the belief that a scientifically sound structure of personality (which they attempted to uncover via the factor-analytic approach) would allow for the reliable prediction and control of behavior. However, empirical evidence gathered by trait theorists did not support using personality indicators as reliable methods of predicting and controlling behavior. It was argued that intervening variables considerably reduced cross-situational consistency in behaviors, so that prediction was difficult (Eysenck 1982). In search of suitable criteria for evaluating trait theories, Cattell (1982) and Eysenck (1982) proposed that constructs should have a demonstrated heritability component. Cattell and Eysenck used behavioral genetic research methods to estimate the extent to which traits are determined by genetic and environmental influences; and found that their constructs were substantially heritable. The condition of heritability has had a lasting influence on both trait and temperament theorists.

Raymond Cattell. According to Cattell, personality research should be the foundation of psychological science (Cattell & Kline, 1977). He believed that viewing psychological process areas (e.g., perception, memory, learning theory, physiological psychology) in isolation provides only a limited glimpse of the whole individual. Instead, he believed that personality is the totality of coexisting variables that influence human behavior. Cattell obtained 4,500 language terms for traits, from a classic study by Allport

and Odbert (1938). He then reduced these terms to a limited number (by first eliminating synonyms, and then using factor analysis) in an attempt to uncover the fundamental structure of human personality. Cattell believed that traits constitute the primary elements of personality. He posited 46 surface traits and 15 source traits. Thirty-six surface traits were obtained by factor analysis, and 10 others were derived from the abnormal psychiatry literature and experimental psychology. He then analyzed the correlations of these 46 surface factors and derived 15 source traits. He also included a 16th factor, intelligence/ability. According to Cattell, source traits are associated with emotional expression, and the sum of these traits is analogous to an individual's temperament. Source traits (as measured by objective methods and revealed through factor analysis) are relatively stable, and describe how a person does what he does (i.e., his general style and tempo). According to Cattell, why a person does what he does also is determined by dynamic motivational traits that tend to fluctuate in response to the environment. Dynamic motivational traits refer to responses to instinctual goals (i.e., ergs); and reactions to people, objects, or social institutions (i.e., sentiments).

Hans Eysenck. Eysenck based his theory on physiology and genetics. However, he also was a confirmed behaviorist who believed in using the principles of learning to make the most of one's genetic endowment. Eysenck viewed personality as growing out of one's genetic inheritance. He placed great importance on individual differences. He believed that researchers working in psychology incorrectly place less value on such differences than do researchers working in other sciences. For example, he wrote that physics and chemistry researchers are aware that one element or alloy does not necessarily behave like another. Therefore, researchers are careful to specify the elements

or alloys with which their experiments or resulting functional equations are concerned. Thus, Eysenck noted the falseness of the assumption that any equation will apply equally to all members of the human species (or to any animal species). Conversely, he believed that idiographic approaches (such as Allport's) went too far.

Eysenck aligned his research with the typological approach. Eysenck's approach differed from Cattell's (and from Guilford's) regarding level of analysis. Cattell and Guilford were interested in trait-level analysis. Eysenck proposed that higher-order aggregates (or types derived from further analysis of the intercorrelations of simple traits) were more empirically robust, and thus more deserving of attention. Moreover, Cattell and Guilford focused on normal personality functioning; whereas Eysenck also was interested in identifying traits underlying pathology. Based on his research, Eysenck (Eysenck & Eysenck, 1985) proposed three factors, extraversion, psychoticism, and neuroticism. The traits that make up extraversion are sociable, lively, active, assertive, sensation-seeking, carefree, dominant, surgent, and venturesome. The traits that make up psychoticism are aggressive, cold, egocentric, impersonal, impulsive, antisocial, unempathic, creative, and tough-minded. The traits that make up neuroticism are anxious, depressed, guilt feelings, low self-esteem, tense, irrational, shy, moody, and emotional.

J.P. Guilford. For more than 25 years, Guilford (Guilford, Zimmerman, & Guilford, 1976) engaged in temperament research using factor analytic methods. Guilford was interested in analysis at the primary trait level, and was unhappy when he found high intercorrelations among primary factors. Guilford believed that factors that measure the greatest number of salient traits would provide more useful information than those that measure a limited number of higher-order traits. The exact number of factors in

Guilford's structure varied as his research progressed and new factors were found. In general, his primary factors closely resemble Eysenck's higher-order factors. In fact, Guilford's factors were used in Eysenck's first studies of extraversion (Cattell & Kline, 1977). Guilford's factors differ from Cattell's mostly because of the factorial methods used to obtain them (Cattell & Gibbons, 1968).

Five-Factor Model. On reviewing earlier factor-analytic work and comparing and contrasting his own research with those of other researchers, Norman (1963) noticed that many researchers who used factor-analytic methods had uncovered five basic factors of personality. Researchers tended to disagree on what these factors assessed. Therefore, these factors were assigned a variety of labels, resulting in confusion and disagreements. In response to this labeling dilemma, Norman assigned roman numerals to the factors.

Newman (1996a) noted that two of Norman's big five factors (I and IV, respectively) are Eysenck's extraversion and neuroticism, with extraversion being the most robust of the five factors. McCrae and Costa added an "openness to experience" factor (one that other researchers had labeled "culture" or "intellect") when developing the original NEO-Personality Inventory (Costa & McCrae, 1985). The "openness to experience" factor became Norman's Factor V. At a conference in 1982, Lewis Goldberg convinced McCrae and Costa to add Factors II and III to their instrument (Newman, 1996a). Factor II was labeled as "agreeableness" and Factor III as "conscientiousness". Thus, the five factor model (FFM) was born.

As McCrae and colleagues hoped, the FFM resulted in cooperative research and cumulative findings for personality psychology. Cross-sectional data from a variety of countries suggest that the factors are transcontextual (McCrae et al., 2000). Within-

culture variation in the factors appears much greater than among-culture variation. Moreover, the factors appear to be shared by nonhuman species. For example, owner's ratings of their pets yielded some of the five factors (Gosling & John, 1998); while zookeeper ratings of chimpanzees yielded all five factors, plus a large dominance factor (King & Figueredo, 1997).

Based on their own research with the FFM (and other relevant research), McCrae et al. (2000) concluded that personality development appears to be an intrinsic process. Except for extreme influences (e.g., trauma) that can have profound effects, developmental timing appears to be largely under genetic control, and environmental influences appear to have relatively little influence on development. In particular, shared experiences (e.g., being reared by the same parents, or attending the same school) show very little effect. In contrast, environmental influences affect the expression of personality. While most behavioral genetic studies attribute most variance in personality to nonshared experiences, much of this variance can be attributed to method and measurement error. For example, Riemann, Angleitnes, and Strelau (1997) obtained data from multi-informant, multi-method sources to estimate method variance. Only 21 to 34% of nonshared variance results from nonshared experiences such as having different peer groups, or teachers; or external biological sources (e.g., prenatal hormonal environment, or central nervous system damage resulting from trauma or disease).

In light of such findings, McCrae and colleagues proposed the traits measured by their personality questionnaires are essentially temperaments (McCrae et al., 2000).

Goodness-of-Fit Theory

Thomas and Chess began their study of temperament in the 1950s (when learning theory and psychoanalytic theories prevailed, and parents were blamed for children's

deviance). Thomas and Chess hypothesized that their contemporaries were assigning too much weight to the role of environmental influences on behavior, and too little weight to constitutional factors. Based on their work with disturbed children and adults, they began to believe that many behaviors traditionally associated with purposive-motivational factors were better viewed as non-motivational behavioral styles (Thomas & Chess, 1977).

Thomas and Chess (1977) described temperament and behavioral styles similarly. Behavioral style refers to how someone behaves, rather than how well, what (i.e., abilities and content), or why (i.e., motivation). However, they distinguished between the two terms. Behavioral style includes stylistic behavioral characteristics that appear both in infancy as well as later in life, while temperament is restricted to those characteristics that are evident in early infancy. They emphasize that their definition of temperament has no implications for etiology or immutability and merely describes the behavioral phenomena. New behaviors that occur later in life may represent older patterns in a new form as well as qualitatively new psychological characteristics. In other words, they believe the developmental process is continuous as well as discontinuous.

Thomas and Chess (1977) characterize development as the complex interplay between children and their environment. The history of this interplay influences future interplay. Because development is complex and without direct causal chains, positive outcomes are possible despite environmental trauma or undesirable genetic endowment.

They borrowed the evolutionary concept of goodness of fit elaborated previously by Henderson (1913) to explain the temperament-environment interactive process. In

adopting the concept, Thomas and Chess adopted the assumption that behavioral style cannot be understood outside of the context of the environments in which it occurs.

Goodness of fit results when there is consonance between the characteristics of a person (e.g., their capacities and behavioral styles) and the characteristics of the environment (e.g., its expectations and demands). Poorness of fit results when there is considerable dissonance between the person and the environment. Goodness of fit does not imply the absence of stress and conflict.

Stress and conflict inevitably occur as people mature into progressively higher levels of functioning and may even serve a constructive role. However, poorness of fit occurs when excessive stress occurs due to a person's inability to meet environmental expectations and demands. Thus, a temperament trait may be influential in some situations and not others as well as at a specific point during one's life and not during others.

Thomas and Chess (1977) developed nine categories of temperament based on an inductive content analysis of parent interview protocols for the infancy periods of the first twenty-two children they studied. Qualitative and factor analytic methods were used to form three behavioral constellations. The categories and constellations are summarized in Table 2-1 and Table 2-2.

Thomas and Chess used undesirable sampling practices in their research. Their first (and most intensively studied) sample, the New York Longitudinal Study (NYLS) sample, was far from representative of the general population. Of the 141 children in the sample, 78% were Jewish, 15% Protestant, and 7% Catholic. Forty percent of the mothers and sixty percent of the fathers had both college education and postgraduate degrees.

Table 2-1. Temperament categories proposed by Thomas and Chess. Adapted with permission from Thomas, A. & Chess, S. (1977). *Temperament and development*. New York: Brunner/Mazel. (pp. 21-22).

| Temperament Category | Definition |
|-----------------------------|--|
| Activity Level | The motor component of a child's functioning and the diurnal proportion of active and inactive periods |
| Rhythmicity | The predictability of biological functions (e.g., hunger, sleep cycles) |
| Approach or Withdrawal | The nature of the initial response (mood expression or motor activity) to a new stimulus |
| Adaptibility | The ease of adapting responses to new or altered situations |
| Threshold of Responsiveness | The intensity level of stimulation needed to evoke a discernible response |
| Intensity of Reaction | The energy level of the response (quality or direction of response not considered) |
| Quality of Mood | The amount of positive or negative affect displayed |
| Distractibility | The degree to which extraneous environmental stimuli interfere with or alter the direction of ongoing behavior |
| Attention Span/Persistence | Two related categories: attention span refers to the time spent on an activity while persistence refers to the continuation of an activity despite obstacles |

Such practices obviously are cause for concern in regards to how these categories and constellations apply to children of parents who are neither Jewish nor well educated. In an attempt to improve the generalizability of their findings by obtaining a sample of contrasting socioeconomic background, Thomas and Chess sampled 95 children of working-class Puerto Rican parents, 86% of whom lived in low-income public housing.

Fortunately, other researchers (Carey, 1970; Carey & McDevitt, 1978) have developed questionnaires based on the dimensions proposed by Thomas and Chess and used sounder sampling practices. Their findings support the generalizability of these dimensions beyond the two samples studied by Thomas and Chess.

Table 2-2. Temperament constellations proposed by Thomas and Chess. Adapted with permission from Thomas, A. & Chess, S. (1977). *Temperament and development*. New York: Brunner/Mazel. (pp. 22-23).

| Behavioral Constellation | Definition |
|--------------------------|--|
| Easy Child | Characterized by regularity, positive approach responses to new stimuli, high adaptability to change, and mild or moderately intense mood which is mostly positive |
| Difficult Child | Characterized by irregularity in biological functions, negative withdrawal to new stimuli, poor adaptability to change, and intense mood expressions which are frequently negative |
| Slow-To-Warm-Up Child | Characterized by a combination of negative responses of mild intensity to new stimuli with slow adaptability after repeated contact. Compared to the Difficult Child constellation, this constellation shows mild intensity of reactions and more regularity of biological functions |

Biological Approaches

Although temperaments are hypothesized to be at the interface of biology and behavior, most research on temperament has focused on behavior rather than biology (Bates & Wachs, 1994). The aim of much of this research has been to establish the stability of temperament traits across the life span. In such research, temperament variables typically have been viewed as either substrates of personality (Goldsmith, Lemery, Aksan, Buss, 2000) or as a subset of personality (Strelau, 1987; Buss & Plomin, 1984). Recent research suggests that temperament is analogous to the big five factors of personality (Angleitner & Ostendorf, 1994; McCrae et al., 2000).

Some researchers began at the biological level and attempted to link biological processes with behavior. Because research needed to map simple behaviors to neural structures required years and temperaments tend to be complex, experiments designed to study nervous system correlates of temperaments have been difficult. As such, the

majority of researchers studying relationships among temperament behaviors and biology have begun at the level of behavior and then attempted to link behaviors with biological factors (e.g., attempts to find relationships among behaviors that are on the introversion-extraversion temperament dimension and brain structures). This method has been called the top-down approach. The top-down approach is considered to be biological in orientation because it places a strong emphasis on biology and views the roots of behavior as biological (Gunnar, 1990). Three well-known top-down approaches include the behavioral-genetic approach, the behavioral inhibition approach, and the reactivity and self-regulation approach.

Behavior-genetic approach. Buss and Plomin (1975) posited four temperaments in their initial theory of temperament: emotionality, activity, sociability, and impulsivity (for which they use the acronym EASI). They established four criteria for deciding what characteristics constitute temperaments: 1) presence in our animal forebears, 2) a strong genetic component, 3) demonstrated stability during the life-span, and 4) adaptive value.

In 1984, they revised their theory (Bluss & Plomin, 1984). They dropped impulsivity as a temperament due to the lack of evidence supporting its heritability. Emotionality was referred to as the intensity of a person's reactions to his or her environment. Arousal was described as the component of emotionality with the most influence on individual differences. The other components of emotionality (i.e., feelings and expression) were believed to be less influential. Activity was described as the total energy output of a person (i.e., the degree to which a person is active or sedentary). Sociability was described as the degree to which the presence of others is or is not preferred over solitude.

Like Thomas and Chess, Buss and Plomin adhere to a behavioral style definition of temperament. According to their later theory (Bluss & Plomin, 1984), temperaments constitute bi-directional variables. Temperaments both influence how others respond to a child and mediate environmental affects on the child. In this theory, temperaments are regarded as substrates of personality that are highly heritable and present early in life.

The heritability of temperaments has been established using quantitative genetics. Plomin and Saudino (1994) note that adoption studies using objective and observational data clearly show a genetic contribution to temperament, but genetics rarely explain more than half the variance in behavior. They believe that temperament traits are regulated by multiple rather than single genes, and that these genes turn on and off during development in order to regulate as well as in response to developmental processes. They propose that molecular genetics will provide undisputable evidence that heritable temperaments exist. Moreover, they propose that molecular and quantitative methods will be used in combination to answer questions about the stability of temperaments, establish links between normal and abnormal behavior, and investigate interactions and correlations between genotype and environment.

Behavioral inhibition approach. Like Buss and Plomin (1984), Kagan considers temperament to be inherited, stable, and manifest during infancy (Kagan, 1989). His work has focused on infants and young children. He takes a biological approach to his research. Thus, he has chosen to focus his study on two temperaments he hypothesized to have empirically observable relationships with biological variables: behavioral approach and behavioral inhibition. Both concepts relate to a child's initial reaction to novelty (e.g, unfamiliar people, objects, and contexts) or challenging situations. He has accumulated

considerable evidence supporting his temperament quality's physiological markers (e.g., heart rate, pupillary dilation, and cortisol levels).

Reactivity and self-regulation approach. Rothbart and Derryberry (1981) defined temperament as constitutionally based individual differences in reactivity and self-regulation. By constitutional, they meant that it is heritable, emerges early in life, and shows stability throughout the life-span. Reactivity referred to individual differences in the excitability or arousal of the nervous system that produce emotional, autonomic, endocrine, behavioral, and other types of responses. Self-regulation referred to the neural and behavioral processes that modulate reactivity. This line of research arose from dissatisfaction with the behavioral style definitions of Thomas and Chess (1977) and Buss and Plomin (1975). Rothbart, Ahadi, and Evans (2000) offer three issues of contention that prompted this alternative to the behavioral style definition. First, temperament traits may not be expressed in all situations. Second, responses may not generalize to all modalities of expression (Rothbart, 1981; Martin, Wisenbaker, & Huttunen, 1994). Third, the study of self-regulation is a useful endeavor absent from behavioral style approaches.

Rothbart and Derryberry (1981) hypothesized substrates of personality could be identified by studying infants and identifying the temperamental components of affect, attention, and action. They found that temperament variables emerge at different ages and that developmental processes require the examination of the structure of temperament at different ages. They found infants to be chiefly reactive, that self-regulatory systems often do not emerge until later in life, and that not all reactive systems are present at birth

(e.g., fear). Moreover, they found that reactive emotional systems can have self-regulatory components (e.g., the behavioral inhibition aspect of fear).

Bottom-up approaches. Bottom-up biological approaches begin with known biological events and attempts to predict behavioral patterns. While simple behavioral processes may be localized to specific structures in the brain, complex psychological processes such as temperaments are more likely to involve interconnected systems and circuits. Thus, attempts to establish brain-behavior relationships when studying temperament are more difficult than when studying less complex phenomena (Steinmetz, 1994). Moreover, at least some individual differences in behavior are likely to result from differences in neural function. Thus, an understanding of brain-behavior relationships on temperament requires researchers to uncover the neural systems involved in temperament and the factors that lead to individual differences in neural function (Steinmetz, 1994).

Nelson (1994) formed hypotheses about relationships that may exist among temperament traits and neural structures. According to his hypotheses, behavioral approach and behavioral inhibition systems are linked with neural structures. The behavioral approach system was described as a system for planning motor movements and directing search behavior toward stimuli. Complex brain functions involve intricate relationships among neural tracts and structures. Nevertheless, an abbreviated explanation of how this system works has been provided by Nelson (1994). The function of this system largely is dictated by two neural structures: the amygdala and the orbitofrontal cortex. Although the amygdala has been implicated in a variety of behavioral functions, within the current discussion it can be thought of as an emotional center. The orbitofrontal cortex is believed to use information from the amygdala and

other sources to determine how a person should behave. If the emotional information received by the orbitofrontal cortex is positive, behavioral approach occurs.

The processing and use of negative emotional information activates the behavioral inhibition system. This system involves inhibiting behaviors or withdrawing from situations. It includes comparator and motor processes (Nelson, 1994). According to Gray's (1991) model of behavioral inhibition, the comparator process has two main functions. First, the subiculum (the part of the hippocampal system involved in memory) receives input from the entorhinal cortex, which receives input from all cortical sensory association areas as well as the amygdala. Second, the Papez circuit (a collection of septohippocampal structures that are involved in processing emotions) appears to facilitate decisions about future events. Motor circuits act on information from the comparator, and the prefrontal cortex feeds information back to the comparator system. Nelson (1994) has summarized information that supports the link between neural circuitry involved in behavioral inhibition and negative affect.

Myers/Briggs Type Theory

The development of the Myers Briggs Type Indicator² (MBTI) can be thought of as a mixture of remarkable foresight and keen observations of human behavior (Quenk, 2000). The MBTI was constructed based on the conclusion that Jung's theory of temperament and his temperament types to describe "healthy personality differences" or "gifts differing". Moreover, the MBTI expanded Jung's theory by adding a new temperament type, judging-perceiving (Myers & McCaulley, 1985; Myers, McCaulley, Quenk, & Hammer, 1998). The judging-perceiving personality types differentiated the

manner in which decisions are made. Persons with a judging temperament make decisions quickly while those with a perceiving temperament prefer postponing them.

From its birth and publication in 1956, interest in the MBTI has grown. It now is used by more than two million people a year to determine normal personality functioning. From 1956 through 1974, the MBTI was popular among only a small number of researchers. However, in 1975, publication of the MBTI was transferred from Educational Testing Service to Consulting Psychologists Press. Movement of the MBTI to the Consulting Psychologist Press made it more accessible to professionals who had the credentials needed to purchase Level B instruments.

Keirsian Theory

According to David Keirsey (1998), environmental influences can not change a person's individual differences. He believes these differences are good and that much can be lost by ignoring or condemning them. To Keirsey, the important differences are innate and develop into a few distinctive patterns. A mature character arises when these traits have developed. Understanding these patterns helps us understand ourselves and others.

Like many typologists, Keirsey posits four temperaments. His work is heavily influenced by the work of Myers and Briggs. After completing the MBTI, results helped him better understand himself and others and ultimately influenced his decision to immerse himself in the study of temperament. He also was influenced by reading Plato's *The Republic*. Plato wrote of four kinds of character corresponding to the four temperaments attributed to Hippocrates. Unlike later typologists such as Jung or Myers, Plato was more concerned with understanding people's roles in the social order than their

² *Myers-Briggs Type Indicator*, *Myers-Briggs*, and *MBTI* are trademarks or registered trademarks of the Myers-Briggs Type Indicator Trust in the United States and other countries.

underlying temperaments. Plato wrote of the iconic (artist), pistic (caretaker), noetic (moralist), and dianoetic (logical investigator) characters. These characters correspond to Hippocrates' temperament types sanguine, melancholic, choleric, and phlegmatic, respectively. Keirsey renamed the iconic character the artisan, the pistic character the guardian, the noetic character the idealist, and the dianoetic character the rational. Keirsey's types correspond to Myers's SP, SJ, NF, and NT types, respectively. According to Keirsey, each temperament has two complementary types and one opposite type based on the use of communication and tools, as visually displayed in Figure 2-1.

Thus, the four temperaments are derived from interweaving communication and tool use. Individual differences result from use of words and tools. Most people are concrete in word usage. People are about equally divided between being cooperative and utilitarian in tool usage.

The attempt to define personality differences by word and tool use sets his view of temperament apart from Myers', which is based on Jung's internal psychological functions. Like Myers, Keirsey expanded the four types into sixteen subtypes. Keirsey's subtypes are called role variants. The artisan role variants are the promoter (ESTP), crafter (ISTP), performer (ESFP), and composer (ISFP). The guardian role variants are the supervisor (ESTJ), inspector (ISTJ), provider (ESFJ), and protector (ISFJ). The idealist role variants are the teacher (ENFJ), counselor (INFJ), champion (ENFP), and healer (INFP). Finally, the rational role variants are the field marshal (ENTJ), the mastermind (INTJ), the inventor (ENTP), and the architect (INTP). Keirsey contrasts the values, interests, self-images, and social roles of the four types. Keirsey also asserted that

the four types differ in their possession of the relative quantity of four kinds of intelligence (i.e., tactical, logistical, diplomatic, and strategic). He notes that, in defining his four “intelligence types”, he is referring to his belief that the kind of intelligent roles displayed by people is determined by their temperament. He is not referring to the degree of skill at that role. In focusing on the practical social role aspect of temperament, Keirsey has done much to explore the implications of temperament in important social matters such as mating, parenting, working, and leading.

| | Abstract Words | Concrete Words |
|----------------------|-------------------------------|-------------------------------|
| Cooperative Tool Use | Abstract NF Cooperator | Concrete SJ Cooperator |
| Utilitarian Tool Use | Abstract NT Utilitarian | Concrete SP Utilitarian |

Figure 2-1. Keirsian types and their derivation. Reprinted with permission from Keirsey, D. (1998). *Please understand me II: Temperament, character, intelligence*. Del Mar, CA: Prometheus Nemesis (p. 29).

SSQ Theory

The Student Styles Questionnaire (SSQ; Oakland, Glutting, & Horton, 1996), like the Myers-Briggs Type Inventory, was founded on the belief that individual differences need not be seen as pejorative, but rather can be used to optimize one's understanding of individual preferences and differences. The SSQ authors, like Myers and Briggs, felt that the knowledge of one's own differences was not sufficient. Instead, this knowledge once acquired should be made functional and used to provide a basis for introspection and self-awareness.

Similar to the MBTI, the SSQ is based on Jungian theory. As previously stated, Jung created his typology for various reasons. One reason was to better equip

psychologists with an understanding of themselves and others. Likewise, teachers equipped with a better understanding of their personal learning styles may be more apt to understand differences in their students and thus design curriculum and provide instruction that are meaningful for them. Moreover, a child involved with learning activities that are personally meaningful and in tune with his or her learning style will put forth more effort in the classroom and enjoy the learning experience. The SSQ, a downward extension of the Myers-Briggs, attempts to provide a link between children's temperament, their education, and other important life events and outcomes. According to SSQ theory, development is shaped by biology, environment, and personal choices or decisions. However, although an environment can be manipulated to optimize learning and children can be taught to make efficient and wise choices, teachers, parents, and others who work with children have little control over a child's biological makeup. Although children can be taught new positive behaviors, the biological makeup of the child impacts the degree he or she is capable of sustaining such an endeavor.

The SSQ is based on normal variation and not pathology or weakness (Oakland et al., 1996). The SSQ can be administered by teachers and then used to help guide decisions regarding instruction and lesson planning. Unfortunately, much of the instruction in the past has placed students in passive learning roles, with the teacher teaching to a class and usually using the same teaching and instructional styles time and time again. However, with a new awareness of their students interests, teachers can create rich and meaningful lessons and form a classroom in which students are active and involved learners because their individual preferences are being respected.

The SSQ was founded on the premise that differences in themselves are strengths and no learning style is superior to another. The impact of a person's temperament depends, in part, on how it enables the person to effectively coexist and interact with his or her environment. Thus, because environments differ and facilitate the display of certain temperament related traits, no temperament type is inherently superior to others. Students taking the SSQ can become aware of areas where they possess talent and aptitude. Knowledge of such can prove highly proactive not only in an educational setting, but also in a metacognitive and personal manner.

The SSQ assesses four bipolar dimensions: extroverted-introverted, practical-imaginative, thinking-feeling, and organized-flexible. In line with SSQ theory, these eight temperament types are referred to as styles. When combined, they form sixteen possible style combinations that are representative of the student's individual differences and preferences.

Extroverted and introverted style preferences. Extroverted and introverted style preferences are related to the source from which a child receives his or her energy. According to Oakland and colleagues (1996), 65% of children prefer the extroverted style and 35% prefer the introverted style. The extroverted child receives his or her energy from others and seeks their company. The extroverted child feels uncomfortable when forced to spend too much time alone. In fact, most extroverted children are uncomfortable with silence.

Children with an extroverted style preference tend to require much attention and communication with their families. They also need more encouragement and praise and thrive in an environment where these are provided. In addition, these children usually get

to know others quickly just as others get to know them quickly. In school, extroverted children are inclined to join groups and enjoy interacting with their peers. They prefer hands-on activities and tend to perform best when long assignments are subdivided into components. They also prefer assignments that can be discharged orally as opposed to employing written expression.

Because extroverted children require more interaction with others, they are more inclined to interrupt their classmates as they work or disrupt the general learning environment. In addition, they are inclined to act on impulse to respond quickly to a task and stop to analyze it after the fact. Finally, they are likely to say the first thing that comes to mind, whether it is a hurtful comment or a compliment.

Children with an introverted style preference draw their energy from within and feel their energy is drained when they spend too much time with others. At home, these children need time to be alone, thus enabling them to derive energy from within, which in turn makes them happier. Although these children tend to enjoy spending time alone, they often have strong attachments to their friends and family. Because they derive energy from being alone, they may be misunderstood by their family, friends, and peers who feel they are too introverted and lacking in social skills. Others may feel that these children prefer to be alone and not participate in activities or family events. Children with an introverted style preference tend to think before speaking. They often are reserved and tend to be cautious in their actions. Because of their tendency to be hesitant, methodical, and less impulsive, their opinions often are well developed and thus respected. These students are more selective of those to whom they get close and to whom their feelings are revealed. In addition, they tend to be sensitive and caring friends because they are

eager to listen to others and make decisions only after considerable thought and introspection.

In school, these students are likely to thrive when their learning and behavioral styles are appreciated. For example, these children can work successfully with others in the classroom if they are placed with compatible peers in small groups or pairs. In addition, because these children seem to possess insight due to their reflective, methodical and logical reasoning, they need opportunities to express their thoughts. Expression can be achieved best by having them present their opinions after they have been given ample time to prepare or they have heard the presentations of others. Unfortunately, because of this trait, these children often get overlooked by teachers who can mistake their reserved nature as uncooperative, unfriendly, and less intelligent.

Practical and imaginative style preferences. Within the school population, about 65% of children prefer a practical style. Students with a practical style are realistic. They attend less to abstractions and more to facts they can use. At home, these children enjoy spending time with their family members and are inclined to possess traditional attitudes towards family life. They find the company of grandparents especially pleasing.

Unfortunately, because these students seem to have strong beliefs and to be drawn to facts, they may avoid peers who are different or possess more imaginative qualities.

Within the classroom, these students enjoy learning sequentially through traditional instructional methods. Many educators who choose to teach in elementary, middle, and high school also have a preference for the practical style and thus exhibit a teaching style consistent with this preference. Students with a practical style preference generally enjoy working with others and are inclined to join groups of peers that share their interests.

They prefer assignments for which there exists an obvious purpose. Abstract concepts are usually rejected by these children as they tend to prefer learning facts that are in turn used for a purpose. In addition, they tend to enjoy activities that are multi-sensory. These children may be uninterested in abstract concepts to the point of rejecting poetry and fictional literature. Finally, because students with a practical style prefer simplicity, they quickly lose interest when given tasks which are time consuming and involve many steps.

About 35% of children prefer an imaginative style. These children tend to display creative instincts. They enjoy activities where new and original ideas are presented and learned through non-traditional means. Children with an imaginative style preference also value the quick acquisition of knowledge and therefore often leaps before truly analyzing the usefulness or applications of a new theory. Thus, children with an imaginative style are inclined to make factual errors. In addition, they become bored more easily and tend to dislike situations or tasks where there is little change. Finally, they tend to enjoy the abstract nature of language and easily interpret metaphors and other types of figurative language.

In the social setting, these children tend to enjoy spending time with other children who share their creative qualities. These children gravitate to other imaginative children, in part, because they may feel rejected by their practical peers and are also inclined to be more accepting of children who are different. Within the family setting, these children generally have strong attachments to siblings and other family members. They may demonstrate their fondness in ways that are out of the ordinary. In the classroom, imaginative children seem to prefer working on projects or assignments in which they are able to use their imagination, learn new skills, and study theories. Unfortunately, because

children with an imaginative style often are attracted to what is novel, they may reject important facts and accept new theories which may lack empirical support. These children also may be sophisticated planners who come up with projects that can become too difficult or elaborate for them to complete.

Thinking and feeling style preferences. Thinking and feeling styles highlight the manner in which individuals make decisions. Unlike the other dimensions, the thinking and feeling styles vary by gender. Approximately 65% of males and 35% of females prefer the thinking style. In contrast, 65% of females and 35% of males have a preference for the feeling style.

Children with a thinking style generally make decisions based on what they believe is fair and logical. They tend to feel little guilt when their decisions affect the feelings of individuals as long as that decision is one that was based on just and logical procedures. In addition, these children are inclined to make decisions only after taking sound and methodical steps to reach a fair conclusion.

Children with a thinking preference tend not to express their feelings. Thus, they often are reserved and tend to avoid situations where emotions and inner feelings are exposed. Children who prefer a thinking style may choose to befriend others who have similar interests. They also prefer to form friendships that have purpose and depth as they are inclined to lose interest in impersonal small talk and gatherings. Disinterest with social talk that seems to have little purpose may limit their social interactions. These children also tend to have deep attachments to their families and tend to be sources of strength during times of need.

In the classroom, students who prefer a thinking style often are competitive and tend to enjoy assignments in which logic is used. They also respond well to praise and recognition for their efforts when it comes in the form of visuals that display their achievements in comparison to that of their peers. Because these students tend to be analytical, they enjoy using computers and reference materials. Consistency generally is important to these students, and they often are skeptical of new information. Unfortunately, because of their tendency to be analytical and skeptical, at times they can be critical of themselves and others and hurt others' feelings while candidly expressing their opinions.

Children with feeling style preferences tend to be diplomats who help to ensure that classrooms are places where everyone is entitled to harmony and respect. To children with this style preference, the feelings of others are valued over logic. Thus, children who prefer a feeling style may make decisions based on how the outcome will affect all parties involved. These students also are more likely to value what they feel strongly about over information that is supported by facts. They often are friendly and talkative, and can be charismatic.

Children who have a preference for a feeling style tend to possess an innate tactfulness that enables them to understand others. In social settings, they tend to enjoy the company of others and are inclined to show empathy and support to friends. Friendship and harmony are a large of part of these children's lives and they tend to enjoy gathering with friends and getting to know them. However, because of their need for harmony, these students may become physically ill during times of conflict and stress in their lives or that of others. Within the family construct, these children tend to be

expressive and affectionate with their siblings and other family members. They tend to feel a great need for belonging and look for such within family and friends.

Within classrooms, these children generally enjoy working with others and respond well to collaborative group assignments. They tend to be interested in learning about the deeds of others and studying history and other subjects related to human nature that promote self-understanding. Unfortunately, these students at times may become too involved in the problems of others to the point of neglecting schoolwork. They also have a difficult time taking sides, thus appearing to be fickle in the eyes of others. Finally, these children at times may become upset to the point of making cutting remarks which can be insensitive to others.

Organized and flexible style preferences. The organized/flexible dimension refers to the propensity to either make decisions promptly or delay them. Approximately equal numbers of children prefer organized and flexible styles.

Children with organized styles generally thrive when they know what to expect. In social settings, students who prefer an organized style have an inclination to be selective about who they befriend, tend to be standard bearers, and may inform others when they find their behavior objectionable. They tend to prefer a traditional and orderly home environment. However, because of their inclination toward that which is orderly and predictable, they may become upset with family members who do not value this type of environment and disrupt the orderly balance.

In classrooms, these children prefer instruction administered in a structured and orderly fashion. They also need to feel that they have some control over the course of their endeavors, including their employment decisions. In addition, they tend to be hard

working, persistent, and finish what they begin. Their need for closure and drawing conclusions may suggest these students are anxious. They tend to put work before play. The desks of organized children tend to be orderly and neatly arranged. They may prefer working on assignments when expectations and grading rubrics are stated clearly. They may respond well to praise that acknowledges their organized and punctual nature. They prefer to have routines within the classroom as well as their personal space in which to organize their materials and belongings. Unfortunately, children with an organized style may be inclined to worry and may hesitate to learn new approaches to doing things. They also tend to be inflexible about their feelings towards issues and may reject those who they feel do not meet their standards.

Children with a flexible style preference enjoy surprises. They generally assimilate easily into different situations and seem to seek experiences in which they have a chance to learn of new ideas and viewpoints. Thus, these children are open minded and tend to be accepting of others. They dislike schedules and prefer to do things at the spur of the moment. Rules tend to be confining. They are inclined to use their wit and charm as gateways that will facilitate the acquisition of new experiences.

Socially, children with flexible style preferences often are fun to be around. They enjoy being with others and tend to be non-judgmental. They tend to feel comfortable in various types of social situations yet may feel uneasy during times of stress. Within their family, they tend to offer amusement. They generally have difficulty following externally imposed rules. Thus, their parents may worry about their carefree nature and lack of following rules. In classroom, they enjoy exploring new ideas and hands-on activities in which they can assemble, disassemble, build, or create objects. They also may prefer to

respond to deadlines that are flexible. Unfortunately, children with flexible style preferences tend to be seen as insensitive by their classmates and may upset them with ill-timed surprises. They also have an inclination to put off assignments and responsibilities in light of other tasks. They also may not contribute equally to group projects or may keep their classmates from doing their work. Finally, they may disillusion others because of their tendency to not keep commitments.

Trait Theory and Type Theory Relationships

Scholars generally view trait theory as more empirically sound than type theory, perhaps because its tenets are easier to test with the common statistical procedures familiar to academic psychology (Newman, 1996b). Differences in the ease with which the two theories can be empirically examined result from differences in test format and statistical procedures. Interestingly, despite theoretical and test format differences, the traits measured by the five factor model are known to correlate with the MBTI types (McCrae & Costa, 1989; Johnson, 1995).

The NEO-PI Extraversion scale correlated at the .7 level with MBTI Extraversion/Introversion dimension, the NEO-PI openness factor correlated at the .7 level with the MBTI sensing/intuitive dimension, the NEO-PI agreeableness factor correlated with the thinking/feeling dimension at the .45 level, and the NEO-PI conscientiousness factor correlated at the .47 level with the judging/perception dimension (McCrae and Costa, 1989). McCrae and Costa suggest that correlations at or above .7 suggest that the scales are essentially equivalent. Although correlations in the .40s are obviously not as strong, they reveal that 16 percent or more of the variance of the agreeableness and conscientiousness factors are in common with those on MBTI dimensions. Using a different type indicator derived from the same item pool used to

create the MBTI, Johnson (1995) replicated the McCrae and Costa study and additionally found a comfort/discomfort dimension on the indicator he used correlated with the NEO-PI neuroticism factor. Despite these empirical relationships, important conceptual differences exist among temperament types and traits (Newman, 1996b; Quenck, 1993).

These differences are summarized in Table 2-3.

Table 2-3. Important conceptual differences between trait and type theories. Adapted with permission from Newman, J. (1996b). Trait theory, type theory and the biological basis of personality. In J. Newman (Ed.), *Measures of the five factor model and psychological type: A major convergence of research and theory*. (pp. 63-80). Gainesville, FL: Center for Applications of Psychological Type (p. 65).

| Trait Theories | Type Theories |
|---|--|
| Individual differences result from quantity of traits possessed | Individual differences result from qualitatively distinct inborn preferences |
| The prominence of a trait is determined by the amount possessed and is indicated by score magnitude | The prominence of a type is determined by the interactions between the categories one is sorted into, score magnitude is used to determine the confidence placed in sorting procedures |
| Normally distributed | Distributions are bimodal and skewed |
| Extreme scores are important for discrimination and may be considered undesirable or diagnostic | Midpoint separating categories are used for discrimination, extreme scores are not important for discrimination and are not considered undesirable or diagnostic |
| Teleological and causal | Teleological |
| Scores may lead to pejorative interpretations | Focus on normal behavior, negative valences not assigned to behaviors |

First, trait theory purports that people differ in the quantity of a trait they possess while type theory purports that people have qualitatively distinct inborn preferences. Second, the prominence of a trait is determined by the amount possessed and is indicated by score magnitude. The prominence of a type is determined by the interactions between the categories into which one is sorted and the score magnitude used to determine the confidence placed in the sorting procedures. Third, the distribution of traits is normal

while the distribution of types is bimodal and skewed. Fourth, in that trait theory involves measuring the quantity of the trait that a person possesses, extreme scores may be considered to indicate too much or too little of a trait and considered undesirable or diagnostic. Type theory uses midpoint-separating categories to make discriminations. Extreme scores are not used to make qualitative judgments are not considered diagnostic. Fifth, trait theory can be interpreted as either teleological or causal (i.e., behavior is caused by traits) while type theory is clearly teleological in that it assumes that type gives purpose to behavior that, in turn, is an expression of type. Finally, trait theory is generally considered to be more pejorative than type theory because type theory is concerned with discovering normal behavior rather than pathology. Consequently, negative valences are not assigned to type behaviors.

Practical Uses of Temperament Concepts

Temperament-based intervention approaches view people as active, selective responders who seek and initiate stimulation and thus play a large role in their own development (Bates, Wachs, & Emde, 1994). Moreover, people are viewed as agents whose social interactions affect how others respond to them. Recommendations based on temperament concepts seldom have empirical validation (Bates et al., 1994). Although such recommendations should remain general and tentative, well-considered applications can help people (Bates et al., 1994).

Temperament approaches almost always involve reframing behavior patterns to enable people to see potential and positive qualities within all traits (Bates et al., 1994). The approaches include predicting possible conflicts to which these traits can lead, constructively dealing with challenging behavior, and learning to use effectively traits that were previously viewed as problematic. Such approaches have the potential to reduce

blame, reduce the use of ineffective approaches to addressing problems, and promote trying novel actions that can resolve conflict between a child and his or her environment.

Effective temperament-based interventions can accomplish important goals. These include increasing skills and knowledge, the quality of social interactions, self-esteem, empathy, and caregiver satisfaction. They also include decreasing caregiver distress and behavior problems. In contrast to such approaches as applied behavior analysis that focus on specific target behaviors that are operationalized so that changes in target behavior can be readily measured, temperament-based interventions must succeed by hitting a largely unspecified target (Bates et al., 1994). As such, measurement of their efficacy is more difficult.

Temperament and adjustment appear to have an empirical relation (Rothbart & Bates, 1998). Certain temperament qualities may predispose children to maladjustment. Temperament information may be useful when formulating diagnoses and treatment planning (Garrison & Earls, 1997). Customized, anticipatory guidance of possible conflicts between children and their environment and how to handle challenging behavior may help prevent psychopathology. For example, children with an extroverted style, face to face counseling appears to be more effective than written guidance. Written guidance may be ineffective and sensitize parents to problems (Cameron et al, 1989, 1991). Strategic family therapists (Fisch et al., 1981; Haley, 1988) suggest that temperament-based approaches can change social systems from a problem focus to one of flexible control that is more accepting of individual differences.

Temperament matching can be achieved by examining the congruence of a child's temperament with significant others in his environment (e.g., parents; Garrison & Earls,

1987). Based on the outcomes of parent guidance interventions, Thomas and Chess (1977) concluded that parental response to temperament of their children tends to be based on congruence with parental goals, standards, and values rather than congruence with temperament. Thus, when developing interventions, comparisons of parents' impressions of their child and their expectations or views of what is considered normal are likely to be more useful than comparisons of child and caregiver temperaments. Lerner's et al. (1982) Dimensions of Temperament Survey enables parents to make such distinctions.

Education

Temperament influences school performance. For example, temperament appears to influence how children approach, engage, and persist with tasks; their classroom behavior; and how teachers respond to them (Keogh, 2003). Interactions among teachers and students are more likely to be positive when teachers interpret children's behavior as temperament-related styles rather than deliberate choices (Keogh, 2003).

Oakland, Glutting, and Horton (1996) presented descriptions of the eight basic temperament styles measured by the SSQ as well as four Keirseian combinations and sixteen style combinations. They recommend curriculum and instructional methods that are likely to best suit specific temperaments and temperament combinations. The intent is to improve student outcomes by capitalizing on strengths and minimizing the consequences of weaknesses. These recommendations offer practical guidance based on intuitive relationships among temperament styles and classroom practices. Some recommendations based on the eight basic temperament styles are detailed in the following paragraphs to illustrate the application of temperament concepts to educational settings.

Children with different learning styles prefer to learn through different types of activities and settings. Children with an extroverted style preference enjoy activities where they are able to express their thoughts and opinions. They thrive when teachers offer constant affirmations and praise. Collaborative learning and peer approaches constitute group practices they enjoy. Children with an extroverted style preference may become bored with activities that limit their chance to interact with peers and share their personal views and ideas.

On the other hand, children with an introverted style tend to prefer to work alone or with one other student. They enjoy activities that require thoughtful planning and metacognitive processes. In other words, these children generally enjoy conducting research on a topic of interest yet resist engaging in activities that require immediate decisions. They thrive in learning environments where they are given a chance to be introspective and analytical. Although these children prefer to work alone, they also can be valuable members in small peer collaborative groups that share their commitments to work.

Children with a practical style preference are likely to shy away from novel situations and innovation. They like to learn information that seems useful to them. They tend to prefer sequential hands-on learning and review of previously learned skills. They are likely to attend closely to detail and make fewer factual errors. Moreover, they are likely to work hard to meet tangible and practical goals.

In contrast, children with an imaginative style preference tend to thrive in situations in which they are presented with novel and innovative information. These children tend to be creative, gravitate towards peers and educators who are different, and seek

knowledge for its own sake. For them, knowledge does not need to have a practical function. They generally enjoy learning new information if it adds to or expands existing views.

Children who prefer a thinking style often communicate briefly and succinctly. They are more likely to be businesslike in their interactions. They prefer competition and tasks that involve the application of logic. They often prefer subjects such as math and science and generally will thrive under learning conditions that embrace methodical thought processes. Similarly, students with thinking preferences also enjoy tasks in which they are asked to analyze and apply information. Most university professors display a thinking style.

In contrast, students who prefer a feeling style are sympathetic and sensitive to other's feelings and thus they are likely to promote harmony and play the role of classroom diplomat. They thrive in positive and cooperative learning environments. They generally enjoy activities that deal with the moral aspects of humanity as well as those that allow them to express their opinions and feelings. Finally, they enjoy working with friends and receiving and providing praise. They often do not respond well to activities that involve competition.

Children who prefer an organized style tend to prefer structured learning environments and tasks. They generally work hard and are reliable. They often enjoy surpassing teachers' expectations and show interest when tasks are factual and objective. They may have difficulty with tasks that have unclear boundaries or expectations or seem to serve no immediate practical purpose.

In contrast, children with a flexible style tend to prefer unstructured learning environments and tasks. They generally benefit from high mobility, task variety, exploration, flexible deadlines, and open-ended assignments. Children with a flexible style are likely to experience school difficulties.

Notably, there has been little empirical investigation of the efficacy of temperament-based curriculum and instructional methods. The research that has been completed indicates that matching temperament type to teaching style does little to improve achievement. Horton and Oakland (1997) found that matching Keirseian temperament combinations to teaching style did not significantly improve students' achievement. Rather, teaching strategies designed to encourage intuitive and feeling qualities such as cooperation, personal application, and identification with the material resulted in significantly higher achievement regardless of temperament.

Vocations

Vocational choices appear to be related to temperament types (Keirsey, 1998; Macdaid, McCaulley, & Kainz, 1986; Oakland, Stafford, Horton, & Glutting, 1996). In addition to vocational preferences, temperament has also been linked to vocationally relevant characteristics such as leadership style and responses to leadership styles (Keirsey, 1998). The MBTI has a long history of use in guiding adult vocational decisions. Thus, the MBTI is used frequently to improve social interactions and ultimately promote efficiency and productivity. Typologists at the Center for Applications of Psychological Type in Gainesville Florida gather data, conduct research, develop empirically-grounded strategies to use MBTI information for the preceding purposes, and offer workshops in the United States and abroad to help promote this information.

The SSQ also can be used to help children and youth identify temperament-related interests that influence vocational interests and choices (Oakland et al, 1996; Oakland et al., 2002). Children think about vocational choices at an early age (i.e., at least at age 8), and their choices are related to temperament type. Strong and uniform relationships between temperament and some vocations can be identified. Notably, the vocational interests of people with flexible styles may be limited. Thus, they may need help to find work that excites them or that they can turn into play (Keirse, 1998; Oakland et al., 1996, Oakland et al., 2002).

Although temperament information can be useful in career planning, it is erroneous to assume all persons with a type of temperament will be happy or successful in the same type of vocation. Individuals with similar temperaments may have divergent interests and talents that are, in turn, influenced by dynamic traits. Moreover, temperaments can be channeled, and opportunities and incentives provided by the social environment can produce adaptations in dynamic personality variables (McCrae et al., 2000) that ultimately can influence job satisfaction.

Considerations in the Generalizability of Temperamental Constructs

Developmental Considerations

The temperaments of young children, especially infants, have been found to be fairly unstable over relatively short intervals of time (Lemery, Goldsmith, Klinnert, & Mrazek, 1999). However, infant temperament has been found to be a modest predictor of adult traits (Wachs, 1994). Self-regulation or control and emotional reactivity have been found to be two of the most important constructs when predicting later behavior (Diener, 2000). As a general principle, individual differences become increasingly stable with age until at least age 30 (McCrae et al., 2000). The functioning of genes is dynamic and

contributes to individual patterns of aging. Therefore, individual differences and traits are not entirely stable. Moreover, even when traits are relatively stable, their expression can change based on contextual factors (McCrae et al, 2000).

McCrae and colleagues (2000) found strong conceptual links among the big five factors and childhood temperaments. They suggest that the same endogenous traits underlie child and adult behavior. In contrast to the proposed similarity of trait types across the life span, they identified age changes in the mean level of traits. For example, between ages 18 to 30 they found neuroticism, extraversion, and openness to experience decrease while agreeableness and conscientiousness increase. These authors could not extend their research into childhood, as it would require a change in instrumentation from the NEO Five Factor Inventory. However, the SSQ has been used to study childhood traits.

Although the NEO is a measure of the five factor model and the SSQ measures Psychological Type, the two approaches show convergence both empirically and conceptually. As such, there should be conceptual links among the developmental trends found with the two measures.

Thayer (1996) identified some developmental trends in the SSQ data. First, she found that the preference for extraversion tends to increase from age eight to the early teen years and begins to level off at age thirteen. Second, younger children generally prefer organized styles and older children generally prefer a flexible style. Third, children develop a stronger preference for feeling styles as they enter their late teen years. No important age trends were found for the practical-imaginative dimension. More recently, Bassett (2004) found that children develop a stronger preference for an imaginative style

between ages 8 and 10, their preferences become more balanced between ages 10 and 15, then a stronger preference for an imaginative style develops again between ages 15 and 17.

Gender Considerations

Although some scholars (Kohnstamm, 1989) believe differences in temperament dimensions are uncommon and the origins and conceptual implications of gender differences are difficult to determine (Maccoby & Jacklin, 1978; Carlson, 2001), gender differences that have been reported. Compared to females, males exhibit higher sensory threshold, less adaptability, less predictability, and less persistence. In addition, males are more likely to be have negative mood and display more difficult temperament patterns.

Using the SSQ temperament dimensions, Thayer (1996) found that boys and girls differ most distinctly on the thinking-feeling dimension. Seventy-two percent of females were found to prefer feeling styles while sixty-four percent of males preferred thinking styles. In addition, boys were found to be somewhat more inclined to both practical styles and flexible styles.

Based on MBTI data, Hammer and Mitchell (1996) report adult males and females differ on the thinking-feeling dimension. Sixty-one percent of females prefer feeling styles while sixty-nine percent of males prefer thinking styles. Males also were more likely to prefer extroverted, intuitive, and flexible styles.

Cultural Considerations

Attempts to define a broad concept such as culture are difficult. Kluckhohn (1954) proposed that "Culture is to society what memory is to individuals." The concept of culture refers to the transmission of shared elements such as norms, values, unstated assumptions, standard operating procedures, and tools (Triandis & Suh, 2002). During

the last few decades, research methodologies for cross-cultural research have been improved (Marsella & Leong, 1995).

Two approaches to address cultural issues have been conceptualized by psychologists conducting cross-national and cross-cultural research: the emic perspective and the etic perspective. The emic perspective emphasizes understanding cultural differences from the viewpoint of those being studied (Marsella & Leong, 1995). Researchers from this perspective attempt to discover indigenous personality constructs prior to adapting constructs from other cultures. In contrast, etic strategies involve the importation and adaptation of constructs into new cultures (Church & Lonner, 1998).

Some proponents of cross-cultural psychology suggest that psychological constructs commonly used in Western psychology have limited applicability in non-Western cultures. For example, Hsu (1985) argued that personality might have limited applicability to non-Western cultures. Others, such as Schweder and Bourne (1984) argue that the very definitions of person and personhood vary across cultures.

Many conclusions about human behavior have been made by Anglos based on their studies of Anglos (Marsella & Leong, 1995). According to these authors, two major errors challenge the validity of many of these studies: the errors of omission and commission. Errors of omission occur when generalizations about human behavior are made without conducting cross-cultural comparisons. Errors of commission occur when researchers include participants from multiple cultures in their research yet conceptualize and conduct research without regard to the points of view of participants from different cultures (e.g., use nonequivalent instruments and assessment methods across cultural groups).

Cross-cultural psychologists have proposed several cultural dimensions to explain the behaviors of cultural populations. Triandis and Suh (2002) identify ecology as an important link to cultural dimensions. Ecological features include isolating landforms (e.g., water and mountains, climate, and the mobility of resources).

Dimensions of culture related to ecology include complexity, tightness, collectivism, and individualism. Complexity refers to the degree of complexity needed to maintain the functioning of a culture. For example, modern information societies rate high on this dimension because maintenance is very complex while societies that are sustained by hunting and gathering rate low on this dimension because maintenance is less complex. Tightness refers to the degree deviation from normative standards are tolerated by persons within a society. Tolerance is more common when population density is low (i.e., there is less opportunity for surveillance) and when people are not highly inter-dependent.

Other commonly cited cultural dimensions include power distance, uncertainty avoidance, and masculinity-femininity. Power distance refers to the extent people accept and expect social hierarchy. Some people may expect and tolerate large inequalities in the distribution of power between individuals, while others accept a smaller power distance. This dimension can be expressed by one's propensity to protest or revolt against establishments that create inequalities. Uncertainty avoidance refers to the extent people find novelty and ambiguity threatening and act to avoid such uncertainty. Uncertainty often is avoided through establishing rules and procedures. Masculinity-femininity refers to the extent people value "masculine" objectives (e.g., money and status) versus "feminine" objectives (e.g., caring for others and quality of life).

Individualism-collectivism is the most frequently researched and cited cultural dimension. Individualism-collectivism refers to the extent that people generalize their sense of responsibility. Individualists think in terms of “I”, where "I" also includes people emotionally close to the individual. Individualists are viewed as inclined to provide for themselves and their immediate family. Collectivists think in terms of “we”, and are inclined to feel responsible for the larger collectivity.

Studies of individualism-collectivism have implications for trait psychology (Church & Lonner, 1998). At the individual level, individualism-collectivism is referred to as idiocentrism-allocentrism. Differences between idiocentrism and allocentrism include views of self as autonomous or connected, whether group or individual goals are prioritized, and the impact of personal attributes or roles and norms on behavior. Church (2000) proposed that temperament traits exist in all cultures, but are more influential on behavior in individualistic societies than in collective societies. He believes situational determinants are more important in collectivist societies.

Fijneman et al. (1996) challenged the belief that individualism-collectivism is a broad value dimension that has a high level of generality. They examined people's expectations to give and seek support and found, contrary to expectations based on individualism-collectivist assumptions, that expectations were similar for people who favored individualist and collectivist orientations. Therefore, they caution against readily accepting broad cultural concepts such as individualism-collectivism to explain behaviors displayed by all people in a culture. Such broad concepts are loosely defined and do not control for confounding variables. Such concepts often are popular for a while and tend to fade in importance with time. Although such concepts can display face validity, their

variability within cultural groups can be substantial. Broad concepts arise in response to a perceived lack of attention to between-culture differences and may fall partially due to the fact that these concepts ignore within-culture differences (e.g., sub-cultures, social status). Thus, cultural differences observed at the country level often have limited predictive value at the individual level (Van de Vijver & Poortinga, 2002). The probability of making errors is rather large when disaggregating psychological variables from the level of a country to an individual.

Piker (1998) has critiqued the conceptual formulations of cultural psychology as disappointing. Cultural psychology proposes that each culture is unique, focuses on the interaction between individuals and culture, and proposes that psychological mechanisms must be studied in situ. Piker notes that cultural psychology either has ignored or denied the inquiry into universal psychological mechanisms.

Nevertheless, cultural psychologists have produced important empirical research in personality and behavior. For example, cultural differences have been found in experience and expression of emotions (Aune & Aune, 1996). The difficult temperament constellation proposed by Thomas and Chess (1977) has been found to be culture dependent (deVries, 1984). deVries examined the survival rates of an African tribe during a severe drought and found that infants with difficult temperaments had a much higher survival rates. He concluded that, although easy temperaments are considered more desirable in industrialized Western culture, a tribe of warriors that values aggressive and assertive behaviors may find the difficult temperament more desirable. Nevertheless, such research suggests that the functional value of temperaments and other psychological

concepts should be interpreted, in part, within a cultural context (Super & Harkness, 1986).

Empirical evidence suggests that some psychological traits are universal. Eysenck (1991) presented evidence that supported the universality of his three superfactors. Similarly, five factor model proponents have proposed that the big five traits transcend differences in language and culture (McCrae & Costa, 1997; McCrae et al., 2000). However, some researchers have argued the existence of factors indigenous to specific cultures. For example, Cheung (1996) has argued that a Chinese "tradition" factor exists and Church, Katigbak, & Reyes (1998) have argued that a "temperamentalness" factor exists in the Filipino culture.

McCrae and colleagues (2000) suggest that these factors may be social attitudes rather than traits. Further research that promotes better understanding of these factors is warranted as their existence challenges the universality of the big five. Moreover, some differences in factors across culture and age-by-culture mean differences in factors have been identified.

The widespread international use of the MBTI provides "extensive and compelling evidence that Jung's theory of personality type is indeed universal" (McCaulley & Moody, 2001, p. 301). However, mean differences in temperaments have been identified cross-nationally using type indicators such as the MBTI and the SSQ (Williams, William, Qisheng, and Xuemei, 1992; Broer & McCarley, 1999; Carlson, 2001; Bassett, 2001; Oakland; 2002). For example, Chinese college students from both mainland China and the People's Republic of China were found to have a higher preference for thinking styles

than U.S. students (Williams, William, Qisheng, and Xuemei, 1992; Broer & McCarley, 1999).

Oakland (2002) compared temperament preferences of children in five countries: Australia, Costa Rica, Greece, People's Republic of China, and the United States. Although all children preferred extroversion, Greek and Costa Rican children exhibited considerably higher preferences while Australian children exhibited discernibly lower preferences for this style. Although children from all five countries preferred thinking styles, the preference for this style was higher in children from Australia and lower in those from the United States. Chinese children preferred practical styles while children from the other countries preferred imaginative styles. Although children from all five countries preferred organized styles, those from Costa Rica exhibited higher preferences and those from Australia lower preferences for this style. According to Oakland, although cross-cultural differences in the magnitude of preferences can be identified, the children from the five countries are more similar than different.

Within the United States, the temperament styles of African American, Hispanic, and White children were compared (Stafford, 1994; Stafford & Oakland, 1996). African Americans and Hispanics were found to be more practical and organized than Whites, with African Americans exhibiting higher preference for thinking than Whites. Despite these differences in mean levels of the types, the structure of the SSQ has been found to be consistent across different cultural groups within the United States (Stafford, 1994; Stafford & Oakland, 1996).

Thus, similar to the cultural differences found in the big five research, cultural differences in psychological types have been found in mean levels of the types, yet the

types themselves seem to remain stable cross-culturally. However, attempts to employ statistical methods (i.e., using factor analytic or confirmatory factor analytic studies) to examine the structure of type indicators cross-nationally have not yet been reported.

Measurement

The measurement of temperament may occur in the context of psychological assessment. Psychological assessment refers to the intensive study of an individual or individuals during which information from multiple sources such as observations, interviews, case history, and test scores are integrated (Anastasi & Urbina, 1997). Information gathered during evaluations is used to make informed decisions on practical matters. Temperament traits have been found to affect behavior in a various settings. Thus, temperament measures provide information that can substantially improve the utility of psychological assessments. Information relevant to developing, using, and interpreting temperament measures is detailed in the following sections.

Psychological Tests

Psychological tests refer to standardized measures used to sample and evaluate specified behavioral domains (American Educational Research Association et al., 1999). Psychological tests can provide information about functioning in a variety of psychological, social, and educational areas. Psychological test uses include facilitating research, describing behavior, identifying talent, certifying attainment of knowledge and other abilities and skills, improving educational and vocational selection, diagnosing disorders, and monitoring change (Oakland, 2004).

An infinite number of observations exists for most behavioral domains. Tests are designed to measure a small, carefully selected sample of an individual's behavior (American Educational Research Association et al., 1999). Temperament consists of

traits that are not directly observable. Rather, temperament traits are latent, inferred from interrelated sets of observations. Inferred characteristics such as temperament historically have been referred to as psychological constructs. Due to confusion and debate about the nature of constructs, the definition of the word construct has been expanded to include the particular concepts or characteristics a test is designed to measure (American Educational Research Association et al., 1999).

Measurement error results from influence on test scores by factors that are not representative of the construct. Error can be both random and systematic (Crocker & Algina, 1986). Random error is derived from purely chance happenings. Random error attenuates both the consistency and accuracy of results. Systematic error is derived from irrelevant characteristics of the test or the person who completed the test. Although systemic error does not affect consistency, it does affect accuracy and therefore the utility of results. Because the measurement process is subject to error, responsible test users must make integrated, evaluative judgments about the validity of inferences based on error estimates.

Validity

Validity is the most fundamental consideration in developing and evaluating tests (American Educational Research Association et al., 1999). The concept refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores (Messick, 1989). Validity is a unified though multi-faceted concept (Cronbach, 1990). Six facets of construct validity have been identified as important in elucidating the central issues in test validity: content, substantive, structural, generalizability, external, and consequential (Messick, 1989).

Content validity refers to the relevance and representativeness of test content. Substantive validity refers to relationships between observed consistencies in test responses and theoretical rationales. Structural validity refers to the fit between the test's scoring structure and the structure of its construct. Generalizability validity involves the identification of sources of error. The extent to which scores and interpretations can be generalized across populations, settings, and tasks is examined and the bounds of test interpretation are established. External validity includes a multitrait-multimethod examination of convergent and discriminant evidence (Campbell & Fiske, 1959) as well as criterion relevance and applied utility (Cronbach & Gleser, 1965). Consequential validity involves the value implications of score interpretation and the actual and potential consequences of test use. Its primary measurement concern is possible adverse consequences that arise from sources of invalidity or test misuse (Messick, 1989).

Decisions about what types of validity evidence are important can be clarified by "...developing a set of propositions that support the proposed interpretation for the particular purpose of testing" (p. 9; American Educational Research Association et al., 1999). After these propositions have been articulated, research must be done to empirically evaluate their soundness. This can be facilitated by examining plausible alternative hypotheses for test interpretation (American Educational Research Association et al., 1999). These hypotheses should address concerns that have been conceived at both a measurement and theory level.

Test Adaptation

A need for assessment tools that can be used with persons from various cultural and ethnic groups increases to the extent societies are culturally diverse. Children from a variety of language, cultural, and ethnic backgrounds attend public schools. Assessment

tools normed on a narrow range of students do not meet the needs of this diverse population. Additionally, increase in the globalization of economic markets, work force mobility, and the cultural heterogeneity of societies contribute to a need for tests that are valid in multiple cultures and languages (Tanzer, 2004).

Tests that are adapted, as opposed to those that are translated, may facilitate their use cross-culturally, cross-nationally, and cross-lingually. Their use saves time and money that otherwise would be devoted to preparing new tests. Until recently, test adaptation and the process of establishing score equivalence across cultures have received little attention (Hambleton, 2003). The transportation of measurement issues across countries, cultures, and languages often has been fraught with faulty practices (Merenda, 2004). Notable errors that have occurred during transportation include the literal translation of items rather their adaptation, failing to re-standardize administration and scoring procedures, using original norms, and failing to exam and confirm the structure of the measurement instrument. In response to such faulty practices, the International Test Commission (ITC) established guidelines for adapting tests across countries, cultures, and languages. The ITC guidelines stress the importance of specifying and justifying valid comparisons between test scores obtained in different languages or cultural groups. Some important issues gleaned from these guidelines and other relevant sources are described below.

Adaptation approaches. The development and implementation of tests for multicultural/multicultural markets generally use one of two methods: simultaneous and successive. The simultaneous test development approach, a newly developed and largely untested approach, calls for tests to be developed simultaneously for use in multiple

languages and cultures. This approach emphasizes the formation of a committee composed of persons from various members from predetermined target groups with which the test will be used. This committee is regarded as an important source of knowledge about the unique needs as well as the linguistic and cultural idiosyncrasies of the target cultures in which the test will be used (Tanzer, 2004).

The successive approach, a far more common practice, involves the development of a test that is validated by several test developers. Once the test has been used successfully for a long period of time, it then is adapted for use in other cultures and/or languages. During the adaptation phase, the original test developer (or a new team who possess skills in psychology, test construction, and psychometrics) reviews the original (i.e., source) test in order to determine the best way to adapt it for the target culture and/or languages.

Translation. The process of translation refers to rendering a test from one language to another. Tests can be translated closely from the source language to the target language. Once the test is translated, the items within the test can be changed to enhance the suitability of the test to varying cultural contexts. Additionally, if the test being adapted is inappropriate in the new culture, a new test can be assembled using a new item pool (van de Vijver, Mylonas, Pavlopouloas, & Georgas, 2003).

Test developers typically use one of two methods to translate tests from the source language to the target language: forward translation and back translation. Using a forward translation method, a translator or group of translators adapt the test from the source language to the target language. Equivalence of the two tests then is assessed by a different group of translators. At times, individuals from the target culture also take the

test. Data from this study may lead to other modifications that could further improve the test.

Although the forward translation method can be effective, it has some drawbacks. At times translators may possess more skill in one language than another and test equivalence is not always judged by individuals with sufficient linguistic, cultural, and psychometric competence. However, the most significant weakness of this method is that the language and other abilities of the bilingual translators may not be equivalent to that of those taking the test. Thus, incorrect assumptions may be made about the skills of the target population who will be taking the test.

The back translation method is best known and most widely used (Hambleton, 2003). Back-translation translators adapt the test from the source language to the target language. The adapted test, which then is in the target language, then is reviewed by a second set of translators and translated back to the source language. Both versions of the source language measure then are compared for equivalence.

The back translation method also has some drawbacks. For example, confusion may occur because comparisons between the source and target versions of the test are done in the source language. In addition, tests that are back-translated and then compared at first may seem to share equivalence; however, upon closer inspection, the grammatical structure and spelling of the source test may have been translated incorrectly to the target language. Therefore, because back translations are done using source language participants, terms that may seem grammatically correct in the source language may have little meaning in the target language.

Many errors can be prevented when persons translating tests are qualified professionals (Hambleton & Patsula, 1999). The belief that knowing two languages makes a person a candidate for being an acceptable translator is a myth. In fact, their use can create significant problems in translations. Thus, translators should be fluent in both languages, knowledgeable of source and target cultures, and knowledgeable of the constructs being studied. For example, a translator who is knowledgeable of English or Spanish and has no knowledge of the construct being assessed will not understand the context in which vocabulary and language will be applied. Words may be translated correctly without capturing their unique meaning within the context of the construct. Analogously, it would be similar to translating the steps to an equation without knowledge of algebra.

Differential item functioning during test adaptation. Differential item functioning assesses whether test takers who have the same total test score have different average item scores or different rates of choosing item options (American Educational Research Association et al., 1999). During test adaptation, differential item functioning can be used to examine for possible item bias. Differential item functioning may help identify terms, expressions, or circumstances that are known to the source population but unknown to the target population. Item bias can occur for many reasons, including item bias due to poor and vague translations and cultural specific items that may have low familiarity in some cultures and high familiarity in others (Tanzer, 2004).

Stafford and Oakland (1996) used differential item functioning to examine the SSQ items for possible bias. These authors examined the responses of three racial-ethnic groups (i.e., African, Anglo, and Hispanic) within the United States. The results indicated

that these racial-ethnic groups used similar response patterns for most items. For the African and Hispanic groups, differential item functioning occurred on only three percent of the items. Differential item functioning occurred on seven percent of the items when the responses of Anglo and Hispanic groups were examined. When the African and Anglo groups were compared, differential item functioning occurred on twenty-five percent of the items, and more than half of those that did are on the Organized-Flexible dimension.

Sampling guidelines. Test developers must be careful in their selection of participants on whom data are collected for use in validity studies. Some recommend selecting participants who share similar background characteristics with target populations (Van De Vijver and Leung, 1996). Such practices reduce error variance due to mistaking sample-specific differences for population differences. Differences in sample characteristics (e.g., motivation) are known to affect responses (Hambleton, 2003). Differences in traditions, norms, values, and how cultures draw meaning from the world may exist within two or more cultural/language groups. Such differences could affect the examinee's views of the assessment tool and performance. Researchers conducting studies across cultures work to ensure the construct being measured by a test is applicable to the target population. An existing equivalent construct in the target group must be present. Certain ideas, expressions, and concepts may exist in the source culture that do not exist in the target culture/cultures. Plausible examples include the indigenous Chinese tradition (Cheung, 1996) and Filipino temperamentalness (Church, Katigbak, & Reyes, 1998) factors proposed by researchers involved in the cross-national study of

personality. Moreover, differing socio-political factors in certain cultural/language groups can affect examinee's scores and should be considered (Hambleton, 2003).

Test format. Responses differ due to the nature of a test's format (Hambleton, 2003). A test format that is both familiar to the target and source populations should be used. Unfamiliarity with test format can be reduced by using detailed, unambiguous instructions that include examples and exercises (van de Vijver & Poortina, 1992).

Test administration. Test administrators are responsible for following established protocols of test administration and maintaining an atmosphere that is conducive to test performance (Glutting, Youngstrom, Oakland, & Marley, 1996). Problems arising from test administrations can be lessened by providing test directions that leave little room for confusion. Test directions should be visual, self-explanatory, and minimize verbal communication (van de Vijver & Poortinga, 1991). Rating scales usually used to measure attitudes seem to cause special problems in test administration (Hambleton, 2000). Test administrators should come from communities being tested, be knowledgeable about their culture, including language and dialects, possess test administration skills, and follow standardized procedures. To this end they should know how test administration can affect validity and reliability of tests, avoid vague communication, and not deviate from test instructions.

Construct equivalence. Construct equivalence should be established when adapting or developing tests for cross-national use. That is, the concept the test is designed to measure should be equivalent or nearly equivalent across different groups (Hambleton, 2003). Construct equivalence subsumes measurement equivalence and theoretical equivalence.

Measurement equivalence refers to the equality of factorial structure across groups. Thus, a test should measure the same number of factors across groups, and patterns of factor loadings should be similar. Moreover, factor loading estimates and measurement error estimates should be similar.

Theoretical equivalence refers to the equality of the theoretical structure. That is, a test should have the same pattern of factor covariance, no significant differences in factor covariance estimates, and no significant differences in factor variance estimates.

Purpose of the Study

Interest in developing tests that can be used internationally has flourished in recent years. As such, attempts to establish the construct equivalence of tests cross-nationally are important. Research investigating the structure of type indicators cross-nationally using factor analytic or confirmatory factor analytic studies has yet to be published. Moreover, evidence of attempts to establish the cross-national construct equivalence of either traits or types for children could not be located.

The purpose of this study is to examine the construct equivalence of a measure of children's temperaments cross-nationally. This research examines whether the SSQ, as used in 8 countries, has etic (cultural-universal) qualities. In other words, does the SSQ measure the same temperament qualities when used with children in various cultures. If the evidence suggests that the SSQ measures the same temperament qualities across countries, children's temperaments, as measured by the SSQ, may assess attributes that transcend cultures. If the evidence suggests otherwise, SSQ interpretations for children outside the United States should be guided in light of differences in the constructs underlying temperament preferences.

Research Questions

Our study addressed the following three questions:

- **Specific aim 1.** Does the SSQ measure the same four bipolar dimensions among children from Australia, China, Costa Rica, Gaza, Nigeria, the Philippines, the United States, and Zimbabwe?
- **Specific aim 2.** If not, what is the structure of the SSQ in these countries?
- **Specific aim 3.** Are the intercorrelations of the dimensions similar across countries?

CHAPTER 3 METHODS

Participants

Participants were 11,784 children from Australia, China, Costa Rica, Gaza, Nigeria, the Philippines, the United States, and Zimbabwe. All of the students attend public school. Data samples from some of the countries were excluded from certain data analyses if they are found to be insufficient in size or quality.

Australian sample. The Australian sample included 369 students (55% female), ranging in age from 9 through 15 (25% in each of the following ages: 9, 11, 13, and 15). They attend public primary and secondary school in the provincial city of Bendigo, in the state of Victoria, Australia. Students were drawn from a cross-section of socio-economic backgrounds.

Chinese sample. The Chinese sample included 400 students (50 % female), ranging in age from 9 through 15 (25% in each of the following ages: 9, 11, 13, and 15), attending public schools in Taiyuan, the provincial capital of Shanxi (the largest city in China). Students were drawn from a cross-section of socio-economic backgrounds.

Costa Rican sample. The Costa Rican sample included 432 students (51% female), ranging in age from 9 through 15 (25% in each of the following ages: 9, 11, 13, and 15) attending both public and private schools. Most participants attend schools that serve middle class families.

Gaza sample. The Gaza sample included 395 students (55% female), ranging in age from 9 through 16 (11% age 9, 14% age 10, 10% age 11, 14% age 12, 11% age 13,

15% age 14, 10% age 15, and 14% age 16) attending public school in Gaza City.

Students were drawn from a cross-section of socio-economic backgrounds.

Nigerian sample. The Nigerian sample consists of 400 students (50% female) ranging in age from 9 through 15 (25% in each of the following age groups: 9, 11, 13, and 15), attending school. Students were drawn from a cross-section of socio-economic backgrounds.

Philippines sample. The Philippine sample consists of 400 students (50% female) ranging in age from 9 through 15 (25% in each of the following age groups: 9, 11, 13, and 15), attending school. Students were drawn from a cross-section of socio-economic backgrounds.

United States sample. The United States participants included 7,902 students (50% female), ranging in age from 8 through 17 (7% age 8, 8% age 9, 12% age 10, 15% age 11, 14% age 12, 15% age 13, 12% age 14, 8% age 15, 5% age 16, 4% age 17). The data were obtained during the standardization of the SSQ using a stratified design representative of the 1990 U.S. Bureau of the Census data.

Zimbabwean sample. The Zimbabwean sample included 492 students (54.5% female), ranging in age from 8 through 14 (25% ages 8 and 9, 25% age 10, 23% age 11, 14% age 12, 12% ages 13 and 14), attending school. Students were drawn from a cross-section of socio-economic backgrounds.

Instrumentation

The Student Styles Questionnaire (SSQ; Oakland, Glutting, & Horton, 1996) is a 69-item self-report temperament scale. The SSQ utilizes a two-option per item, forced-choice response format. Items have a third grade readability level in English. It was designed to be used with children ages 8 through 17.

The SSQ consists of four empirically-derived primary dimensions. These dimensions are measured on bipolar scales, and their theoretical structure is based on Jungian typology theory: extroversion-introversion, practical-imaginative, thinking-feeling, organized-flexible. These dimensions yield eight styles that in turn yield 16 possible combinations of four preferred styles based on the Myers-Briggs model.

The SSQ scores appear to be fairly stable over time, with test-retest reliabilities ranging from .67 to .80. An average test-retest reliability coefficient of .74 was obtained over a 9 month period, using a Fisher's z transformation applied to coefficients from the four scales. The use of dichotomous data restricts variance at the item level and precludes a true estimate of internal consistency. Measures of internal consistency coefficients generally over-estimate a test's reliability when used with dichotomous data and thus were not used. Factor analytic data show that the internal structure of the test is stable and consistent with the theoretical framework of the instrument for persons who differ by age, gender, and race/ethnicity (Oakland, Glutting, and Stafford, 1996; Stafford & Oakland, 1996).

Convergent validity studies reveal that MBTI dimensions correlate well with three of the four SSQ scales for a group of 12 to 17 year-old students. In particular, a joint canonical correlation analysis revealed strong correlations between the extroverted-introverted, thinking-feeling, and practical-imaginative scales from the SSQ and MBTI. The aforementioned relationships are reported as canonical loadings, and range in magnitude from .84 to .96. Canonical loadings equal to or greater than an absolute value of .80 are considered substantial (Oakland, Glutting, & Horton, 1996).

Divergent validity studies indicate that the SSQ constructs are distinct from academic achievement and intelligence. Many of these correlations are less than .1. None of these correlations exceed .3.

The suitability of the SSQ for use with students in Australia, Costa Rica, Gaza, Nigeria, The People's Republic of China, the Philippines, and Zimbabwe was reviewed in light of guidelines established for the translation of tests by the International Test Commission (Hambleton, 2003). For example, issues pertaining to the suitability of item content, format and response style were reviewed by Dr. Oakland's collaborators in each of these countries.

Upon review, the following collaborators determined that the SSQ was suitable for use with students in Australia, Costa Rica, Gaza, Nigeria, The People's Republic of China, the Philippines, and Zimbabwe, respectively: Dr. Michael Faulkner, professor and Head of Department, Institute for Education, La Trobe University, Bendigo, Australia; Ana Lorena Mata who served as a school psychologist in Los Angeles for almost 20 years prior to her return to her native Costa Rica; Dr. Mohammad Adnan Alghorani, Assistant Professor of Psychology, The University of Oman; Dr. Andrew Mogaji, professor of psychology, University of Lagos, Nigeria; Dr. Lu Li, a professor at the Shanxi Medical University, Shanxi Taiyuan, People's Republic of China; Carmelo Callueng, professor of education at the Philippines National University, Manila, Philippines; and Dr. Elias Mpofu, professor of rehabilitation psychology at Pennsylvania State University. All five are fluent in English.

Procedure

The SSQ is an English language test. The language used in this test was found to be suitable for use in Australia, Nigeria, The Philippines, and Zimbabwe. However,

translations were needed prior to its use in Costa Rica, Gaza, and the People's Republic of China. The SSQ was adapted using back-translation sequential methods in these three countries. The SSQ in English (i.e., SSQ-E1) was translated into Spanish (i.e., SSQ-S) for use in Costa Rica by Ana Lorena Mata, into Arabic (i.e., SSQ-A) by Dr. Algoorgani for use in Gaza, and into Cantonese (i.e., SSQ-C) for use in the People's Republic of China by Dr. Lu Li. Colleagues fluent in both English and their native language translated the SSQ-S, SSQ-A, and SSQ-C back into English (SSQ-E2). SSQ-E1 and SSQ-E2 were compared to determine their equivalence. Adjustments in language were made in SSQ-S, SSQ-A, and SSQ-C as needed.

The SSQ was administered to groups of students in their schools. The students were asked to read and mark items appropriately, consistent with the test's directions. Students used either an answer form or marked the items directly on the test booklet. Psychologists were available to offer assistance and answer questions.

Data Analysis Procedures

Raw scores were used for data analyses. Raw scores are more appropriate than normalized scores in this type of research because the distribution of raw scores on the extroverted-introverted, practical-imaginative, thinking-feeling, and organized-flexible dimensions may not be normal. Moreover, normalized and weighted scores would distort the actual score distribution.

The SSQ uses dichotomous items. Data analysis of dichotomously-scored items is considerably more complicated than data from items that are continuously scored (e.g., a Likert scale). The Jungian theoretical structure of the test requires a force-choice format. Jungian theory holds that people may be defensive about their preferences and thus may report their true preferences inaccurately if a Likert type system is used. The forced-

choice format is thought to be less susceptible to this type of response bias. A response format that uses a broader scale may sacrifice important clinical information pertaining to preferences in favor of providing more suitable empirical data for examination of construct validity.

Unfortunately, when using dichotomous variables, attempts to compute Pearson correlation coefficients for all relationships may not be possible due to restricted variance. Item parceling can be used when working with dichotomous data (Oakland, Glutting, & Horton, 1996). Item parceling involves combining items into groups, thus producing larger variances.

However, item parceling procedures have disadvantages. First, as parcels must be chosen, these procedures are more subjective than analyzing individual items. In addition, an examination of the factor loadings of individual items is not possible using this approach. The test authors (Oakland, Glutting, & Horton, 1996, page 183-184) indicate that an approach involving the repeated factoring of parcels, opening a different parcel during each successive factoring, can be used to avoid problems associated with both the factoring of individual dichotomous items and the factoring of item parcels.

In addition, tetrachoric correlations can be computed when using dichotomous data. These correlations are computed using a formula for binary data. Binary data may violate the distributional assumptions of the maximum likelihood method. However, the use of tetrachoric correlations allows for the retention of binary data because observed covariances were corrected prior to being analyzed in a CFA. In contrast to item parceling, this approach eliminates the need for a priori grouping of items.

Several methods exist for engaging in the analysis of test structure across countries: factor analysis, multidimensional scaling, cluster analysis, and covariance structure analysis (van de Vijver & Leung, 1996). Each is described below.

Factor Analysis

Factor analysis involves the examination of covariation among observed variables in order to identify the most robust unobservable (latent) factors that contribute to test performance (Byrne, 1998). The technique also involves examining and interpreting how and to what extent the observed variables are related to the latent factors. The extent of such relationships is represented by factor loadings. When more than one latent variable is identified, hierarchical analyses of these latent variables can be performed to determine whether a higher-order latent factor exists. An exploratory factor analytic approach is used when no prior knowledge exists as to whether the observable variables actually measure the intended latent construct (Byrne, 1998).

A strategy for applying factor analysis in cross-cultural and cross-national analyses of test structure is discussed by van de Vijver and Leung (1996). This strategy also can be applied to cross-national analyses of test structure. First, they recommend determining if the same number of factors is evident within different countries, with factors selected based on amount of variance each factor explains. In reference to the SSQ, the underlying dimensions are bipolar. Thus, a two stage factor-analytic process is required to differentiate the components of the dimensions. The dimensions of the test are thought to be essentially independent. When factors are thought to be independent, principal axes factor analysis using varimax rotations has been recommended in order to rotate the structure orthogonally (Merenda, 1997).

Second, if the same number of factors is found in the different groups being compared, then test items are examined to determine if the items correlate with the same underlying factors in each group. Third, the equality of the factor loadings can be determined by using a procedure that begins with target rotations. This procedure involves arbitrarily choosing one set of factor loadings as a target and rotating the axes of the other factor loading matrices so that agreement between the sets is optimized.

Then, the similarity of the loadings is evaluated during sequential comparisons through a coefficient of agreement (e.g., Tucker's coefficient of agreement; Tucker, 1951). The powerfulness of this approach is questionable (van de Vijver and Leung, 1996). In addition, the influence of nonequivalent items is not reflected in the coefficient of agreement.

Multidimensional Scaling and Cluster Analysis

Multidimensional scaling (MDS) or cluster analysis can be used to compare the structure of cross-cultural and cross-national data sets. MDS refers to a class of techniques used to uncover the underlying structure of data (Kruskal & Wish, 1978). MDS procedures use proximities as input. Proximities refer to quantitative values assigned to datum that collectively indicate similarities and differences in the data.

In reference to the SSQ, proximities could be derived from either intercorrelations of or squared distances between items. The output of MDS procedures consists of spatial representations that display the geometric configurations of data points along dimensions. Researchers can vary the number of dimensions in the output, thus allowing them to identify the most interpretable and theoretically parsimonious structure. A goodness-of-fit (or more aptly, a “badness-of-fit”) measure known as stress is used to determine the number of dimensions needed to best describe the structure. Stress is defined as the

square root of a normalized residual sum of squares. The larger the stress values, the poorer the fit.

As with factor analysis, a rotation problem occurs when attempting to compare the results from different countries. The same approach used during factor analytic procedures (i.e., performing a target rotation and using an index of agreement) can be used to examine similarity of the solutions (van de Vijver and Leung, 1996). However, there are no known published examples of research in which target rotations have been applied during the cross-cultural analysis of structure.

Cluster analysis refers to a set of multivariate statistical procedures used to form homogenous clusters of data points derived from similarities within data sets (Aldenderfer & Bashfield, 1984). In reference to the present study, clusters obtained during analyses would represent temperament dimensions or other factors that might be measured by the SSQ. Although possibly suitable for cross-cultural and cross-national research, cluster analysis rarely has been implemented with these forms of data.

Covariance Structure Analysis Techniques

The data could be submitted to covariance structure analysis techniques in order to examine cross-national differences in test structure. These techniques attempt to explain relationships among a larger set of observed variables in terms of a smaller number of unobserved variables. Relationships among the observed variables are characterized by covariances that are contained in a covariance matrix (Long, 1983b).

The theory behind this approach typically holds that the unobserved variables are “causal” processes assumed to produce the pattern of relationships (i.e., structure) among multiple observed variables (Byrne, 1998). The causal processes are represented by a series of regression equations. This multivariate technique allows for the simultaneous

testing of many variables, thus allowing complex structural models to be hypothesized and tested statistically in order to determine the extent of their consistency with the data. Various goodness-of-fit measures are used (Byrne, 1998). Using confirmatory factor analysis (CFA; Joreskog, 1969), the structure of the SSQ can be specified a priori and either confirmed or rejected in favor of competing models based on empirical findings.

Data Analysis Procedures used for the Present Study

In the present study, the extent to which the SSQ measures the four primary dimensions based on Jungian typology theory was examined. The structural equation modeling (SEM) technique of CFA is commonly used and often preferred by contemporary researchers who conduct empirical tests of hypotheses about tests with latent structures (Keith, 1997), and therefore was selected as the most appropriate method for the present research purposes. Tetrachoric correlation matrixes and covariance matrixes were analyzed using AMOS 5.0 (Arbuckle, 2003) computer software following the method of maximum likelihood, which assumes a multivariate normal distribution.

CFA was used to examine the SSQ structure for each country separately and in simultaneous multi-country comparisons. CFA allows for the empirical testing of theories and hypotheses, which researchers accomplish by setting substantively motivated constraints (Long, 1983a). The following constraints were imposed:

- Intercorrelations of latent factors
- Specifying which observed variables are affected by which latent factors
- Constraining the error variances of observed variables to zero when necessary due to identification problems

The four bipolar dimension model, as well as plausible competing models (i.e., a model with three bipolar dimensions and a hierarchical model) were tested for each

country. These alternative models were chosen based on empirical and theoretical rationales. A model with three bipolar dimensions was chosen to test the hypothesis that the data better conform to a simpler, more parsimonious model. Had this hypothesis been accepted, a two bipolar dimension would have been examined. A five dimension model also could have been examined. If, for example, the SSQ contained items that appeared to be feasible measures of neuroticism, a neuroticism factor could have been added. However, this was not done because a rationale for assigning SSQ items into a fifth bipolar dimension was not identified. A hierarchical model was chosen to test the hypothesis that the data better conform to a model containing a second-order factor in addition to four bipolar dimensions.

Model fit for individual countries was evaluated using fit indexes that indicate how a specified model of interest compares with an independence model in which all observed variables are uncorrelated. Fit indexes are used to decide whether the null hypothesis of a close fit of the model to the data should be rejected. They may be interpreted more aptly as poorness of fit indices because they do not confirm models, only disconfirm them. There are dozens of fit indexes described in the SEM literature. Because these indexes reflect different facets of model fit, multiple indexes are presented. The fit indices used to evaluate individual models evaluated included the following:

- Chi-square test
- Ratio of chi-square to degrees of freedom
- Goodness-of-Fit Index (GFI)
- Comparative Fit Indices (CFI)
- Tucker-Lewis Index (TLI)
- Root Mean Square Error of Approximation (RMSEA)

The chi-square test has degrees of freedom equal to the difference between the number of observations and the number of parameters. This statistic is interpreted as the

difference between the model specified by the researcher and a just-identified version of it. Small and nonsignificant values are desired. However, the chi-square test typically is not used as a significance test because, when sample sizes are large, the statistic may be significant even though differences between the observed and model-implied covariances are small. The ratio of chi-square to degrees of freedom provides a measure of fit that is less sensitive to sample size. Although there is no definitive value used to establish acceptability, a ratio of less than 3 is frequently suggested (Kline, 1998).

The GFI is considered to be more standardized and less sensitive to sample size than the chi-square. The GFI is interpreted as the proportion of observed covariance explained by the model-implied covariance. The CFI is an index with an analogous rationale, it indicates the proportion in the improvement of fit of the model specified by the researcher relative to an independence model (i.e., the observed variables are assumed to be uncorrelated). Although the Tucker-Lewis Index (TLI) is similar to the CFI, it includes a correction for model complexity. Acceptable fit for the GFI, CFI, and TLI indexes is considered to be indicated by values exceeding .90 (Kline, 1998).

The Root Mean Square Error of Approximation (RMSEA) is a standardized summary of the average covariance residuals. Covariance residuals are the difference between the observed and model-implied covariances. The smaller the value, the better the fit. Values less than .05 generally are considered indicative of acceptable fit (Browne & Cudeck, 1993).

Fit indexes show only the average fit of a model. Some parts may fit well while other parts may fit poorly. Standardized regression weights were examined in order to examine the model more closely and to better understand its theoretical meaningfulness.

When comparing competing models or making multi-group comparisons, fit indices useful for making such comparisons are emphasized. For example, the change in chi-squared, along with degrees of freedom and associated probability, were used to make such comparisons. The Akaike Information Criterion also was used. This criterion compares models that may not necessarily be nested. Lower criterion values indicate better fit. In addition, the GFI, CFI, TLI, and RMSEA were examined. Moreover, invariance in factor loadings, factor correlations, (i.e., factor structure), and measurement residuals (i.e., error in the prediction of latent factors) were tested in multi-country comparisons.

Notably, the SSQ utilizes item weights to evaluate the strength of children's preferences. That is, some items are weighted more heavily than others in the scoring system. Item weighting was derived empirically by analyzing the standardization sample. Such weights are subject to capitalization on chance variation within the particular sample. The present study is concerned with internal structure of the SSQ across samples from a variety of countries. Therefore, item weights were not considered in the CFA models tested.

CHAPTER 4 RESULTS

Five sets of confirmatory factor analyses (CFA) were needed to test the hypotheses in this research. The first set involved testing the structure of the SSQ using a first-order CFA model with the sixty-nine items loading on four bipolar temperaments, as specified in Appendix B of the SSQ manual. This appendix delineates item scales and weights. The model tested is shown in Figure 4-1. A statistical solution could not be calculated for most countries because many items did not have sufficient variance. Restricted item variance poses a statistical problem commonly encountered when using binary data. Due to this problem, a second set of analyses was carried out.

The second set involved testing the validity of a first-order CFA model of item parcels the test authors derived from factor analytic and heuristic considerations. More specifically, a factor analysis of correlations among dichotomous item scores provided factor loadings, and items were combined into four- to five-item parcels based on the similarity of these loadings. This model is shown in Figure 4-2. Some estimates of parcel variance were found to be unreasonably large. This problem was addressed by constraining variance parameters to zero as needed. However, the unreasonably large estimates raise questions about the correctness of the model. Merely constraining problematic parameters does not address these issues fully. Therefore, a third set of CFA was needed.

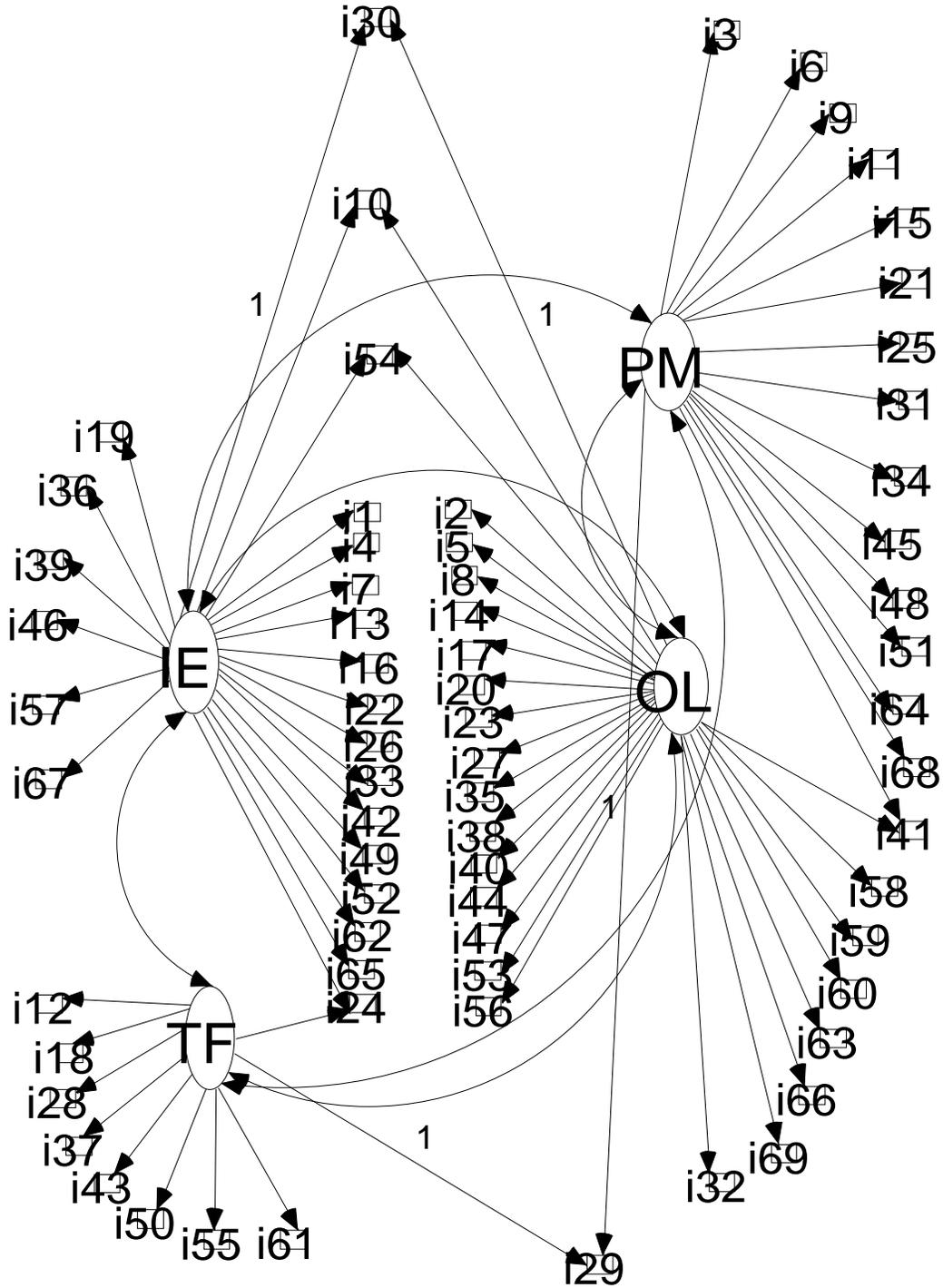


Figure 4-1. Sixty-nine item Student Style's Questionnaire model.

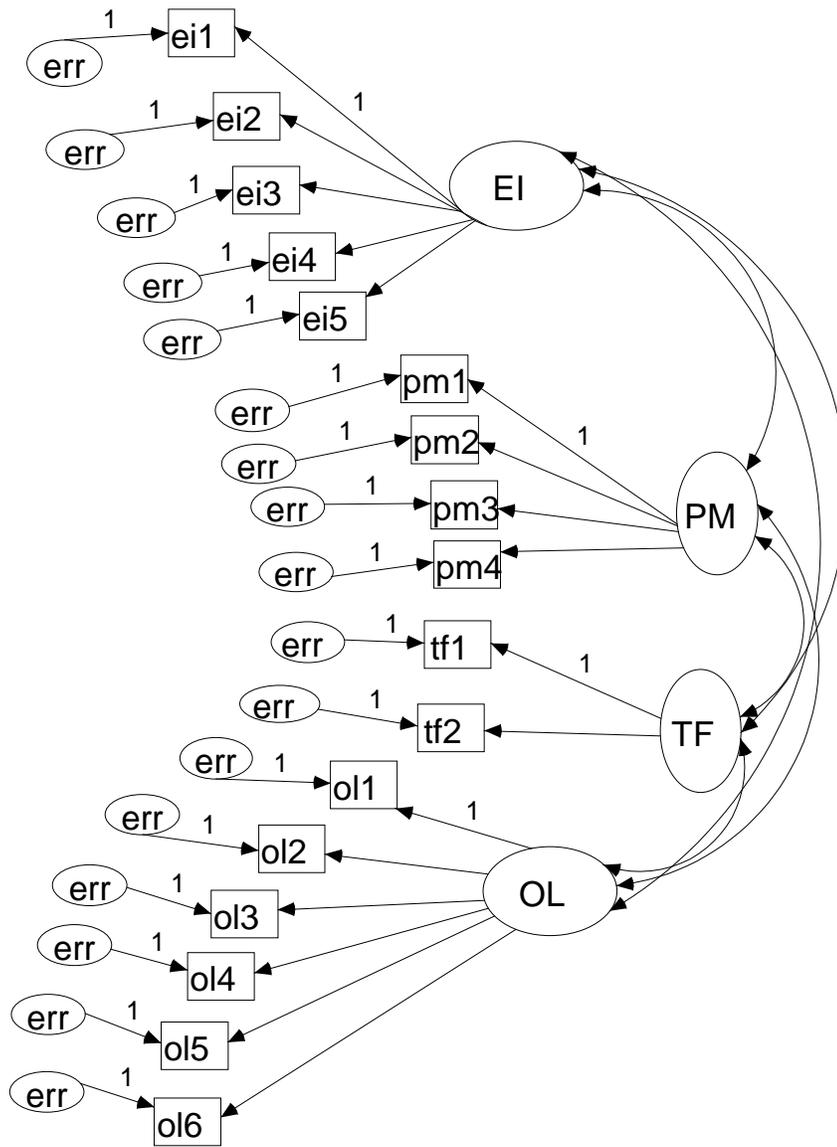


Figure 4-2. Parcel model.

The third set of CFA involved testing the validity of a first-order CFA model of modified item parcels. The modifications were guided by the assumption that model fit was reduced by unspecified relationships between certain items. The modification process is described in greater detail later along with the results obtained from using this set of CFA. The model is shown in Figure 4-3.

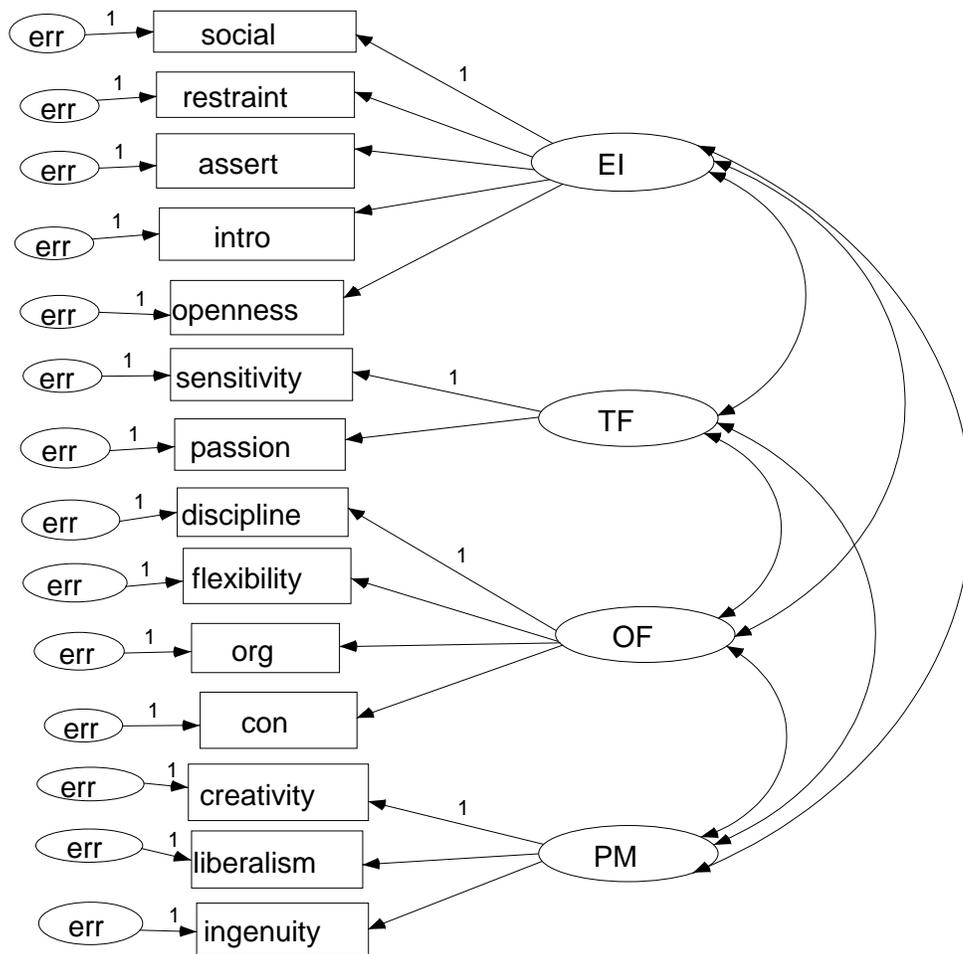


Figure 4-3. Modified model.

Fourth, a three bipolar temperament model (Figure 4-4) was examined. Fifth, a hierarchical CFA model (Figure 4-5) was examined. The fourth and fifth sets of CFA were completed to examine plausible rival hypotheses that either a three-factor model or a hierarchical model best represents the data. An examination of competing hypotheses is an integral part of strong programs of construct validation (Keith & Kranzler, 1999).

When possible during each of the five sets of CFA, analyses involved examining model fit for individual countries as well as multi-group comparisons. The rationale that guided the progression of these five sets of analyses is shown in Figure 4-6.

The first data analysis approach attempted was to conduct a first-order CFA of the 69 SSQ items and their proposed four-factor relationship. Specifically, this approach tests the hypothesis that children's temperament is a multidimensional construct composed of the four SSQ bipolar styles. The model (Figure 4-1) specified relationships between individual items and the four bipolar styles based on the listing of items and their related scales in Appendix B of the SSQ manual (Oakland, Glutting, & Horton, 1996). In order to enhance the visual clarity of this figure, item error parameters are not represented. However, these parameters were included in CFA analyses.

The latent content variables (i.e., the bipolar styles) were allowed to correlate. Factor loadings and measurement errors were set free to vary for most observed variables. In that unobserved variables have no definite metric scale, some parameters had to be fixed to satisfy this scaling requisite. Specifically, one factor loading in each set of factor loadings thought to measure the same latent variable was constrained to 1.0. Likewise, in that error is an unobserved variable and therefore has no definite metric scale, the regression weights associated with error were constrained to 1.0 for each item.

SSQ item responses were used as input, and matrixes of tetrachoric correlations were computed using a formula for binary data. The obtained matrixes of tetrachoric correlations were used as input for CFA analysis. The model specified is overidentified. This means the ability to calculate a unique estimate of every parameter is theoretically possible. However, overidentification is a necessary but insufficient condition to resolve the identification problem (Byrne, 2001).

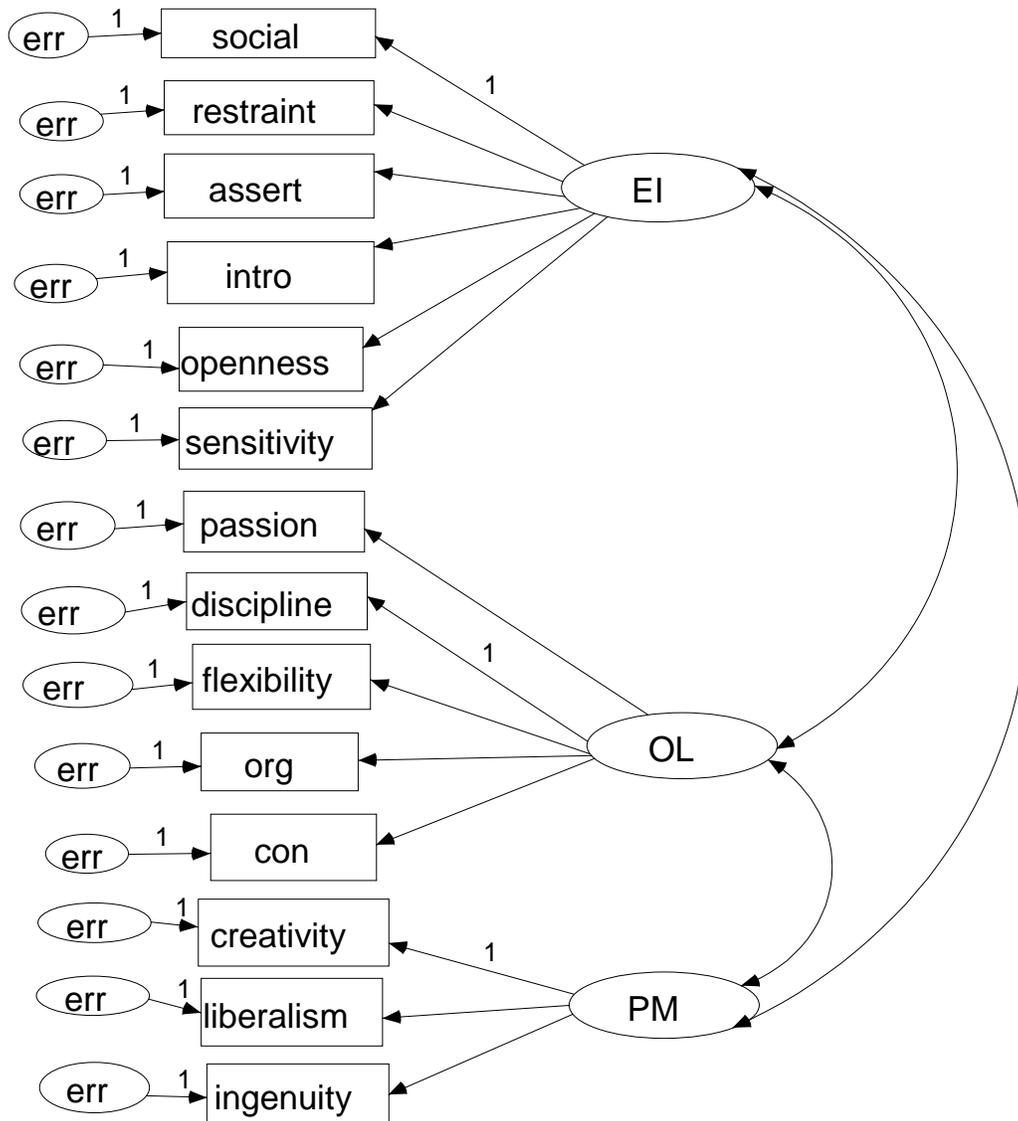


Figure 4-4. Three-factor model

The model in Figure 4-1 was tested empirically for the United States sample only. The resulting fit indices suggest that the model provides inadequate fit to the US data. Modification indexes indicate that model fit could be greatly improved by allowing some measurement error terms to correlate. However, allowing measurement error terms to

correlate without substantive reason would result in a theoretically meaningless model that does not answer any questions related to this study.

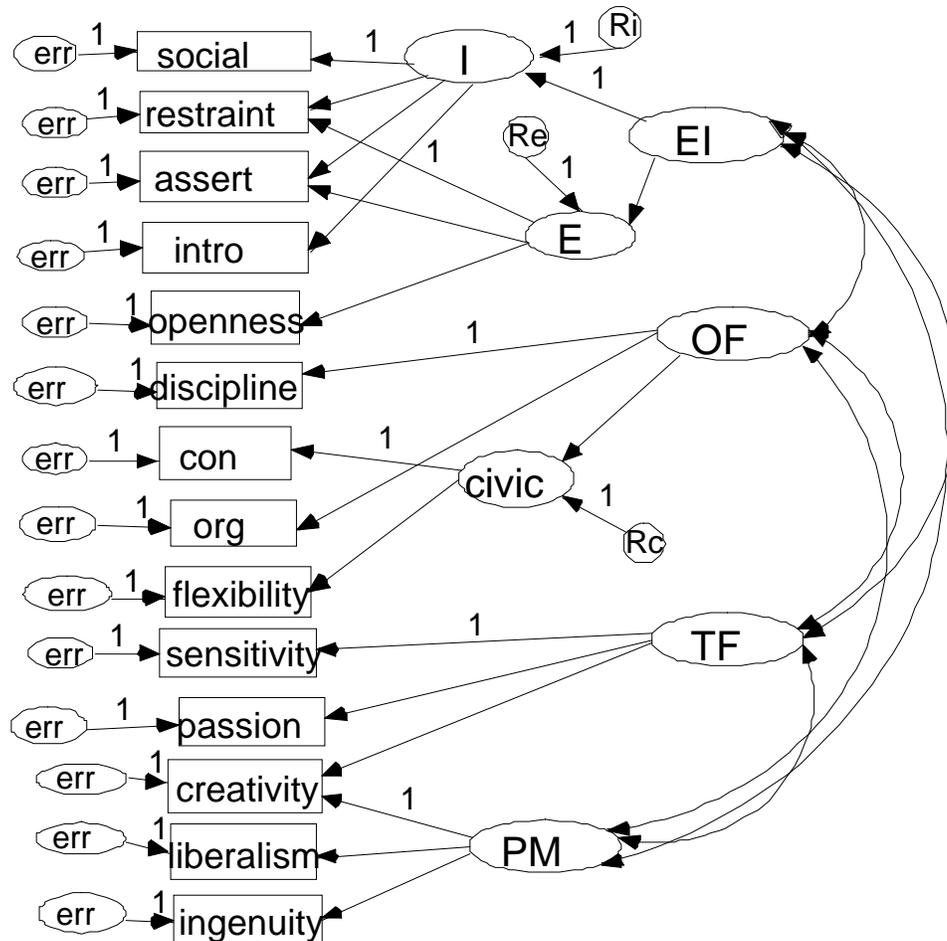


Figure 4-5. Hierarchical model.

A matrix is nonpositive definite when it contains an out-of-bounds correlation, resulting in the failure of certain mathematical operations with the matrix (Kline, 1998). The use of tetrachoric correlations and multicollinearity (i.e., highly correlated items) are known causes of nonpositive definiteness.

Inspection of correlation matrixes revealed a few out-of-bounds correlations. Thus, nonpositive definiteness of sample correlation matrixes prohibited the evaluation of this CFA model with data from other countries. Also, many correlations were found to be

close to zero. An attempt to identify specific problematic items that may have contributed to this problem was unsuccessful. Likewise, examination of data from multiple countries suggests that out-of-bounds correlations obtained when correlating specific item pairs tend to be specific to that country. The use of item parcels in subsequent analyses eliminated both the need for tetrachoric correlations and multicollinearity. In summary, the first set of analyses provided little useful information for answering the research questions proposed in this study.

The parcels used in the second set of analyses were obtained using factor analytic and heuristic methods. A review of the SSQ test manual led to the belief that factor loadings obtained from exploratory factor analyses and content validity were primary considerations used to assign items into parcels. Some test items were placed into multiple parcels and used to measure two different bipolar temperament dimensions. These parcels, as delineated in the SSQ manual, are presented in Appendix A. The model is shown in Figure 4-2.

This model specified factor loadings between the four bipolar styles and the parcels purported to measure them. The latent content variables (i.e., the bipolar styles) were allowed to correlate. With the exception of reference variables, factor loadings were set free to vary. Measurement errors were set free to vary, while the regression weights associated with these terms were constrained to 1.0.

The fit indexes for the eight countries in this study are detailed in Table 4-1. Standardized regression weights and correlations between the bipolar styles are presented in Appendix B. Notably, certain error parameters were not identified in some samples.

1st Set

The sixty-nine item Student Style's Questionnaire model provides inadequate fit to the US data and could not be used to evaluate data from other countries.

**2nd Set**

The parcel model provides good fit for all countries except Gaza. Multi-group comparison indicates reasonably good fit; however, the chi-square difference between the unconstrained model and a model with factor loadings constrained to be equal across groups is statistically significant and indicative of noninvariance. Also, a number of estimates of parcel variance had to be constrained due to their magnitude.

**3rd Set**

Post hoc re-specification of item parcels was undertaken so that fewer constraints (i.e., estimates of parcel variance) had to be relaxed. The development of this modified model, that was guided by modification indexes and consideration of item content, allows for better examination of construct relevant variance (i.e., the degree to which scores measure processes that are relevant to the constructs of interest and not to irrelevant constructs or measurement error). This model provides good fit for all countries; however, the variance of measurement error terms for discipline and creativity parcels had to be constrained to zero in order to achieve identification for the Zimbabwe sample. In addition, for the Australian sample, the correlation between EI and OF and the correlation between EI and PM had to be constrained to zero to achieve identification. Simultaneous examination indicates reasonably good fit; however, the chi-square difference between the unconstrained model and a model with factor loadings constrained to be equal across groups is statistically significant and indicative of noninvariance.

**4th and 5th Sets**

The three-factor and hierarchical models were examined to test plausible rival hypotheses. The three-factor model was identified for only three countries (i.e., Australia, China, and Costa Rica) and provides a less adequate fit than competing models. The hierarchical model provides the best fit for Australia, China, Gaza, Nigeria, and USA data.

Figure 4-6. Rationale for analyses.

This problem was addressed by constraining these variances to zero. The measurement error for parcel tf1 was constrained to zero for Australia. The measurement

errors for parcel tf1 and ei5 were constrained to zero for Gaza. The measurement errors for parcels ei3 and pm1 were constrained to zero for Nigeria and the Philippines.

Fit indexes suggest that the model provides a poor fit for the Gaza data. These indexes provide mixed support of model fit for other countries. For all countries except the United States and Gaza, the ratio of chi-square to degrees of freedom is indicative of adequate model fit. The chi-square value for the United States sample appears to be strongly influenced by sample size. The fit indexes less sensitive to sample size are more in line with those obtained for the other countries. The GFI index suggests adequate fit for all countries except Gaza. The RMSEA index indicates reasonable fit for all countries except Gaza. In contrast, the CFI and TLI indexes are indicative of questionable fit and suggest that at least some part of the model is misspecified.

Table 4-1. Comparison of fit for the parcel model.

| Country | X^2 | df | X^2/df | GFI | CFI | TLI | RMSEA |
|-------------|---------|-----|----------|-----|-----|-----|-------|
| Australia | 270.89 | 114 | 2.38 | .92 | .65 | .58 | .06 |
| China | 334.90 | 113 | 2.96 | .91 | .65 | .58 | .07 |
| Costa Rica | 261.33 | 113 | 2.31 | .93 | .78 | .73 | .06 |
| Gaza | 537.667 | 115 | 4.68 | .86 | .42 | .31 | .10 |
| Nigeria | 305.672 | 115 | 2.66 | .92 | .45 | .35 | .06 |
| Philippines | 239.12 | 115 | 2.08 | .94 | .64 | .57 | .05 |
| USA | 3267.72 | 113 | 28.92 | .95 | .62 | .55 | .06 |
| Zimbabwe | 284.24 | 113 | 2.52 | .94 | .57 | .46 | .06 |

Results for the multi-country analysis are presented in Table 4-2. Factor loadings, structural covariances, and most measurement residuals were constrained to be equal. Measurement errors for pm1 and tf1 were constrained to zero. The results are mixed. The GFI and RMSEA indexes suggest adequate model fit. In contrast, the difference in chi-square between the unconstrained model and a model with factor loadings constrained to

be equal across groups is statistically significant and indicative of noninvariance (i.e., the structure varies across groups). Moreover, AIC, CFI, and TLI indexes suggest noninvariance. Thus, the results suggest that some of the equality constraints do not hold across countries.

Table 4-2. Fit of the parcel model, multi-country analysis.

| Model | χ^2 | df | χ^2/df | AIC | GFI | CFI | TLI | RMSEA |
|------------------------|----------|------|--------------------|------|-----|-----|-----|-------|
| Unconstrained | 6052 | 920 | 6.58 | 6660 | .94 | .57 | .49 | .02 |
| Measurement Weights | 6808 | 1011 | 6.73 | 7234 | .93 | .51 | .48 | .02 |
| Structural Covariances | 7124 | 1081 | 6.60 | 7410 | .93 | .49 | .49 | .02 |
| Measurement Residuals | 7624 | 1186 | 6.43 | 7700 | .92 | .46 | .50 | .02 |

The covariance structure analysis model considers interscale correlations to derive from scale loadings on correlated factors (Long, 1983b). An examination of the magnitude of parcel factor loading shows some variance across countries. This is not surprising in that these parcels were formed, in part, on item factor loadings obtained using exploratory factor analysis. This analysis has the statistical danger of producing patterns of item loadings that are largely due to sampling error (Agresti & Finley, 1997). An examination of the modification indexes showed that modifying the model to allow measurement errors to correlate would significantly improve model fit. In other words, the parcels share variance not accounted for by the four bipolar dimensions.

Although the content (i.e., item) similarity of the parcels is obvious, finding meaningful distinctions between them is difficult without an understanding of the meaning of the parcels. Allowing for correlated measurement errors makes the model meaningless.

For example, correlated measurement errors could signify the existence of unobserved, exogenous variables not specified in the model. Moreover, if the measurement error terms of parcels believed to measure the same construct are correlated, their relationship could derive entirely from measurement error.

The results of any confirmatory factor analysis depend on the variables that are entered. Some SSQ items appear to be more similar in content to certain items that have been specified as measuring the same temperament than to other items specified as measuring the same temperament. In other words, groups of items appear to measure qualities that underlie temperaments.

Post hoc specification of these qualities was undertaken in an attempt to examine construct relevant variance (i.e., the degree to which scores measure processes that are relevant to the constructs of interest and not to irrelevant constructs or measurement error) more closely. Although specifying these qualities does not change how SSQ items covary with the intended structure of the theoretical domain, such specification allows for the improvement of overall model fit and more interpretable results. Notably, the combination of items into parcels that allow for greater improvement in fit and theoretically meaningful results may be possible.

The approach used to form this particular combination involved first examining modification indexes for the model in Figure 4-1. Specifically, modification indexes for allowing item pairs to correlate were examined. Many large and statistically significant modification indexes were found. Next, content of items from pairs with large modification indexes was subjectively examined to identify similarities. These

similarities were specified in a modified model as qualities that underlie the bipolar temperaments. The item content of the specified qualities is shown in Appendix C.

The modified model is shown in the Figure 4-3. The fit indexes for the eight countries in this study are detailed in Table 4-3. The standardized regression weights, latent factor correlations, and fit indices for the respective countries are presented in Appendix D.

Table 4-3. Comparison of fit for the modified model.

| Country | X^2 | df | X^2/df | GFI | CFI | TLI | RMSEA |
|-------------|---------|----|----------|-----|-----|-----|-------|
| Australia | 187.65 | 73 | 2.57 | .94 | .49 | .36 | .07 |
| China | 204.72 | 71 | 2.88 | .93 | .53 | .40 | .07 |
| Costa Rica | 99.62 | 71 | 1.40 | .97 | .81 | .75 | .03 |
| Gaza | 168.88 | 71 | 2.38 | .94 | .56 | .43 | .06 |
| Nigeria | 110.91 | 71 | 1.56 | .96 | .61 | .49 | .04 |
| Philippines | 94.33 | 71 | 1.33 | .97 | .79 | .73 | .03 |
| USA | 1157.44 | 71 | 16.30 | .98 | .62 | .51 | .04 |
| Zimbabwe | 120.00 | 73 | 1.64 | .97 | .50 | .38 | .04 |

Notably, for the Zimbabwe sample, the variance of the measurement error terms for discipline and creativity had to be constrained to zero in order to achieve identification. In addition, for the Australian sample, the correlation between EI and OF and the correlation between EI and PM had to be constrained to zero to achieve identification. As with the last set of analyses, the chi-square value for the United States was found to be greatly influenced by the large sample sizes. The ratio of chi-square to degrees of freedom, the GFI index, and the RMSEA generally support adequate model fit for all countries. In contrast, the CFI and TLI indexes are indicative of questionable fit.

Results for the multi-country analysis are presented in Table 4-4. Factor loadings, structural covariances, and measurement residuals were constrained to be equal.

Zimbabwe was excluded from analyses because of concerns about the quality of the data related to the inflated variance of the measurement error terms for the discipline and creativity parcels.

Table 4-4. Fit of the modified model, multi- country analysis.

| Model | X ² | df | X ² / df | AIC | GFI | CFI | TLI | RMSEA |
|------------------------|----------------|-----|------------------------|------|-----|-----|-----|-------|
| Unconstrained | 2685 | 667 | 4.03 | 2821 | .96 | .49 | .51 | .02 |
| Measurement Weights | 3029 | 677 | 4.48 | 3145 | .96 | .40 | .44 | .02 |
| Structural Covariances | 2960 | 687 | 4.31 | 3056 | .96 | .42 | .46 | .02 |
| Measurement Residuals | 3087 | 701 | 4.40 | 3155 | .96 | .39 | .45 | .02 |

The obtained results are mixed. The GFI and RMSEA indexes suggest adequate model fit. In contrast, the difference in chi-square value between the unconstrained model and a model with factor loadings constrained to be equal across groups is statistically significant and indicative of between country differences. Likewise, this chi-square difference is indicative of noninvariance with latent factor intercorrelations and measurement residuals. Moreover, AIC, CFI, and TLI indexes suggest noninvariance. Thus, the results suggest that at least some of the equality constraints do not hold across countries.

Model trimming and building were used to develop some alternative models. A three-factor model is presented in Figure 4-4. The thinking-feeling dimension is excluded from this model because it is measured by fewer items than the other dimensions and, as indicated by previous factor analyses presented in the manual, its eigenvalue was found to be small compared to some other dimensions.

A hierarchical model is presented in Figure 4-5. This model specifies latent introversion and extroversion variables that underlie the high-order extroversion-introversion dimension. This model also specifies what appears to be a latent variable that underlies the organized-flexible dimension. Notably, the existence of these latent variables is compatible with the four bipolar dimensions interpretation. The existence of underlying variables does not diminish the importance of higher-order variables, in this case the four bipolar dimensions.

A three-factor model was identified for only three countries. The fit indexes for this model are presented in Table 4-5. The hierarchical model was identified for five countries. The fit indexes for this model are presented in Table 4-6.

The three-factor, hierarchical, parcel and modified models were compared (Table 4-7). The results show that the hierarchical model to be best for Australia, China, Gaza, Nigeria, and the USA. The modified model provides the best fit for Costa Rica, the Philippines, and Zimbabwe. The results provide support for the four bipolar temperament interpretation of the SSQ.

Table 4-5. Fit for identified three-factor models.

| Country | X^2 | df | X^2/df | GFI | CFI | TLI | RMSEA | AIC |
|------------|--------|----|----------|-----|-----|-----|-------|--------|
| Australia | 159.03 | 74 | 2.12 | .94 | .62 | .53 | .06 | 221.03 |
| China | 260.65 | 74 | 3.52 | .91 | .34 | .19 | .08 | 322.65 |
| Costa Rica | 101.08 | 74 | 1.37 | .97 | .82 | .78 | .03 | 163.08 |

Table 4-6. Fit for identified hierarchical models.

| Country | X ² | df | X ² / df | GFI | CFI | TLI | RMSEA | AIC |
|-----------|----------------|----|------------------------|-----|-----|-----|-------|--------|
| Australia | 90.30 | 65 | 1.39 | .97 | .89 | .84 | .03 | 170.26 |
| China | 144.81 | 65 | 2.23 | .95 | .72 | .61 | .06 | 224.81 |
| Gaza | 109.63 | 65 | 1.69 | .96 | .80 | .72 | .04 | 189.63 |
| Nigeria | 83.05 | 65 | 1.28 | .97 | .82 | .75 | .03 | 163.05 |
| USA | 525.55 | 65 | 8.09 | .99 | .83 | .77 | .03 | 605.55 |

Table 4-7. Comparison of competing models.

| Country | X ² | df | Drop in X ² | Change in df | AIC |
|-----------------|----------------|-----|------------------------|--------------|---------|
| Australia | | | | | |
| 1. Modified | 187.65 | 73 | | | 251.65 |
| 2. Three-factor | 159.03 | 74 | 28.62 | +1 | 221.03 |
| 3. Hierarchical | 90.30 | 65 | 97.35 | -8 | 170.26 |
| China | | | | | |
| 1. Parcel | 334.90 | 113 | | | 414.90 |
| 2. Three-factor | 260.65 | 74 | 74.25 | -39 | 322.65 |
| 3. Modified | 204.72 | 71 | 130.18 | -42 | 272.72 |
| Costa Rica | | | | | |
| 1. Parcel | 261.33 | 113 | | | 341.33 |
| 2. Three-factor | 101.08 | 74 | 160.25 | -39 | 163.08 |
| 3. Modified | 99.64 | 71 | 161.69 | -42 | 167.62 |
| Gaza | | | | | |
| 1. Modified | 168.88 | 71 | | | 236.88 |
| 2. Hierarchical | 109.63 | 65 | 59.25 | -6 | 189.63 |
| Nigeria | | | | | |
| 1. Modified | 110.91 | 71 | | | 178.91 |
| 2. Hierarchical | 83.05 | 65 | 27.86 | -6 | 163.05 |
| United States | | | | | |
| 1. Parcel | 3267.72 | 113 | | | 3347.72 |
| 2. Modified | 1157.44 | 71 | 2110.28 | -42 | 1125.44 |
| 3. Hierarchical | 525.55 | 65 | 2742.17 | -48 | 605.55 |

CHAPTER 5 DISCUSSION

The purpose of this study was to examine the construct equivalence of the Student Style's Questionnaire (SSQ) cross-nationally. The following three questions were addressed:

- **Specific aim 1.** Does the SSQ measure the same four bipolar dimensions among children from Australia, China, Costa Rica, Gaza, Nigeria, the Philippines, the United States, and Zimbabwe?
- **Specific aim 2.** If not, what is the structure of the SSQ in these countries?
- **Specific aim 3.** Are the intercorrelations of the dimensions similar across countries?

Five sets of analyses were needed to address the questions described above. The rationale used to guide these analyses are described in Figure 7. The results of this study clearly support the conclusion that the SSQ measures the same four bipolar dimensions or styles among children from Australia, China, Costa Rica, Gaza, Nigeria, the Philippines, The United States, and Zimbabwe.

Notably, the parcel model did not provide good fit for the data provided by children from Gaza. This finding may reflect difficulties with translation, as the Arabic language used by these children uses a different alphabet and language structure. This finding also may relate to religious and cultural beliefs, as these are the only group of Muslim children who participated in the study. Moreover, the data were obtained during Intafada, a five year period of conflict between the Palestinians and Israelis. Trauma related to Intafada may have influenced the responses of these children.

In addition, the measurement error terms for the discipline and creativity parcels were found to be inflated when testing the modified model with data provided by children from Zimbabwe. This finding may have occurred because the children from this country have been subjected to more poverty than the children from the other countries in this study. Moreover, although these children receive instruction in English, their first language is a local language. This may have hindered their comprehension of test items.

Some of the underlying qualities that constitute the bipolar styles display some between-country differentiation. For example, the hierarchical model, as shown in Figure 6, specifies latent introversion and extroversion variables underlying the high-order extroversion-introversion dimension and also specifies a latent variable underlying the organized-flexible dimension. The hierarchical model provides the best fit for some countries but not for others, suggesting that between-country differences exist in the qualities underlying the extroversion-introversion and organized-flexible dimensions. However, the four bipolar styles represent the broadest, most general dimensions of temperament. Despite possible between-country differences with some of the underlying qualities that characterize these dimensions, the four bipolar styles appear to capture the rich complexity of children's temperament. The four bipolar style model is not incompatible with differentiated (e.g., hierarchical) models that focus on more specific qualities.

Also, the intercorrelations of the dimensions display some between-country differences. SSQ theory holds that the dimensions are uncorrelated. In general, weak to moderate intercorrelations were found. However, the Extroverted-Introverted and Thinking-Feeling dimensions correlate for Australian and Zimbabwean children. The

Practical-Imaginative and Thinking-Feeling dimensions correlate for Philippine children. The Thinking-Feeling and Organized-Flexible dimensions correlate for children from Gaza. Nevertheless, despite these significant intercorrelations, a four dimension model provides the best fit for the data obtained from all eight countries (i.e., Australia, China, Costa Rica, Gaza, , Nigeria, the Philippines, the United States, and Zimbabwe).

Implications

Cross-National Implications

Recent research suggests that temperament is analogous to the big five factors of personality (Angleitner & Ostendorf, 1994; McCrae et al., 2000). Five factor model proponents have proposed that the big five traits transcend differences in language and culture (McCrae & Costa, 1997; McCrae et al., 2000). The results of the present study are congruent with the conclusion that temperaments transcend differences in language and culture. The present study extends these results through its use of confirmatory rather than exploratory factor analytic techniques. The use of confirmatory techniques makes the present study a more rigorous and direct evaluation of temperament structure than those provided by McCrae and colleagues.

Methodological Implications

The present study provides a cross-national examination of the structure of a forced-choice measure of children's temperament. Accomplishing this endeavor yielded some methodological implications that may be useful to researchers interested in conducting similar research.

Confirmatory factor analysis. Although several methods (e.g., exploratory factor analysis, multidimensional scaling, cluster analysis, and confirmatory factor analysis) exist for engaging in the analysis of test structure across countries (van de Vijver &

Leung, 1996), confirmatory factor analysis has several advantages over alternative methods.

First, confirmatory factor analysis (CFA) requires a priori specification of model parameters (Byrne, 1998). A priori specification relies on theoretical guidance and reduces the probability of making incorrect decisions about structure. Second, researchers using CFA are able to simultaneously perform analyses for multiple countries. The present study tested invariance in factor loadings, factor correlations, (i.e., factor structure), and measurement residuals. Researchers conducting multi-group analyses also could test for invariance of intercepts and means. Third, assuming some differences are found, CFA may be used to examine each country and empirically test hypotheses. Finally, researchers using CFA may directly test and compare the relative fit of plausible competing models (Byrne, 1998). The examination of rival hypotheses is a hallmark of strong validation programs (Keith & Kranzler, 1999).

Open constructs. The SSQ bipolar styles are conceptualized as open constructs (Pap, 1953). Open constructs obtain meaning through theoretical variables. Thus, attempts to specify observable behaviors by which the constructs can be evaluated are difficult (Millon, 1991). The development of the SSQ was driven by Jung's theory of temperament as augmented by Myers and Briggs. Although the four bipolar styles that correspond to the Jungian temperaments are the concepts of interest, the present confirmatory factor analyses suggest that underlying variables exist. This finding indicates that failure to specify underlying variables in the confirmatory factor model may yield erroneous conclusions about test structure.

Dichotomous data. The SSQ uses dichotomous items. Dichotomous variables restrict variance and may violate the distributional assumptions of the maximum likelihood method typically used in confirmatory factor analyses. To address this problem, tetrachoric correlations were computed and used in confirmatory factor analyses. Tetrachoric correlations are computed using a formula for binary data that corrects observed covariances.

Scholars generally have viewed trait theory as more empirically sound than type theory, perhaps because its tenets are easier to test with the common statistical procedures familiar to academic psychology (Newman, 1995b). Tetrachoric correlations and confirmatory factor analysis may be valuable techniques for researchers who use dichotomous data, such as those interested in testing the tenets of type theory.

Test Validity Implications

Relative to many psychological tests (e.g., published tests of intelligence and academic achievement), the Myers-Briggs Type Indicator (MBTI) and the SSQ are less amenable to empirical examination with commonly used statistical procedures. Personality assessment in general has been critiqued as lagging behind other areas of psychological testing such as the measurement of intelligence and achievement (Martin, 1988). The construct validity of most existing personality measures has not been substantiated with strong programs of construct validation. Such programs emphasize the importance of theoretical guidance during test development and validation (Benson, 1998) as well as the recognition of plausible rival hypotheses (Keith & Kranzler, 1999). Testing practices supported by strong programs of construct validity adhere to the standards established jointly by the American Educational Research Association et al. (1999). SSQ development was guided by theory. The current study examines plausible

alternative hypotheses about test structure and contributes to the process of substantiating the measure with a strong program of construct validation. The results of the present study provide additional construct validity support for the SSQ. Past studies of SSQ structure have utilized factor analytic procedures (e.g., Oakland, Glutting, & Horton, 1996; Stafford, 1994; Stafford & Oakland, 1996). The results of the confirmatory factor analyses in the present study are consistent with the results of factor analytic studies. Both types of studies have yielded four bipolar dimensions.

The SSQ was developed as an extension of the MBTI. Although weaknesses are evident with personality measures in general, the psychometric properties and overall merit of the MBTI specifically have been vociferously criticized (e.g., Pittenger, 1993). However, recent confirmatory factor analytic research (Bess, Harvey, & Swartz, 2003) supports the construct validity of the MBTI. The MBTI is used frequently in applied organizational and assessment settings for such purposes as career-counseling, self-development, team building, and highly litigious purposes such as employee selection and placement (Bess, Harvey, & Swartz, 2003). Although the SSQ was developed as an extension of the MBTI, its popularity with test users in educational settings does not parallel the popularity of the MBTI in business settings.

The SSQ is similar to the MBTI and can be used for many of the same purposes. Theoretical considerations were prominent in the development of both instruments and both appear to measure the constructs they were intended to measure. Given the immense popularity and long history of the MBTI, educators appear to be underutilizing the SSQ.

Moreover, the results of the present study suggest that the SSQ may be appropriate for use cross-nationally. Thus, the test could be normed and marketed internationally.

Limitations

One weakness of the present study pertains to sample sizes. The United States sample was large. Although the sample sizes obtained from the other countries in this study would be considered large when judged by other studies, they are considered marginally large enough for the use of confirmatory factor analysis. Confidence in the accuracy of results increases as sample size increases.

Another weakness is found in the best fitting model obtained following post hoc assignment of items into qualities. A priori specifications are empirically superior to post hoc specifications. This fact does not challenge the validity of the underlying theory, as the temperament types are conceptualized as collections of behavioral consistencies. The results do suggest that assigning items into sub-temperament qualities may be an empirically useful consideration when designing self-report measures of temperament.

Directions for Future Research

Future studies might involve administering other measures of temperament and personality (e.g., the NEO-PI) simultaneously with the SSQ. Confirmatory factor analytic techniques could be used to empirically test the convergent and discriminate validity of the temperament and personality measures administered to participants. Convergent and discriminate validity refers to the degree constructs correlate with and differ from variables in accordance with theoretical predictions (Campbell & Fiske, 1959). Such studies would be useful in extending knowledge pertaining to the structure of and relationship between temperament and personality.

Watson (2000) has shown that affective (i.e., mood) traits are strongly related to four of the Big Five dimensions (all but Openness) of personality. By extrapolation, the SSQ styles likely have strong affective components. Future research is needed to test this

hypothesis. Moreover, useful information might be obtained by exploring relationships between the SSQ styles and other possible lower-order qualities (e.g., sociability or conscientiousness).

The results of the present study suggest the SSQ styles may correlate strongly in some countries (e.g., Australia, Gaza, the Philippines, and Zimbabwe). Future research is needed to examine these interrelations in order to determine if and why strong interrelationships exist.

Some SSQ items may function differently in different countries. Therefore, future research is needed to examine the SSQ at the item level. For example, some translated items may not have retained their original meaning. Moreover, it may be useful to examine cultural differences at the level of item type, as recent research suggests that certain types of items may be more prone to bias than others (Osterlind, Miao, Sheng, & Chia, 2004).

Future research also might focus on the application of the SSQ to solve applied assessment problems. For example, research could examine the use of SSQ data to guide pedagogical strategies. SSQ data also could be used for team building purposes (e.g., for students who collaborate in the classroom as well as in extracurricular activities such as sports). Many such studies have been undertaken by researchers interested in the MBTI (Bess, Harvey, & Swartz, 2003). Although the SSQ is similar to the MBTI and inferences from its scores are likely useful for many of the same purposes, such claims should be subjected to empirical examination.

APPENDIX A
LISTING OF PARCEL ITEMS

Table A-1. Parcel items for EI (extroverted-introverted).

| Parcel | Items in Parcel |
|--------|--------------------|
| ei 1 | 1, 26, 33, 52, 62 |
| ei 2 | 4, 19, 36, 49 |
| ei 3 | 7, 24, 46, 60 |
| ei 4 | 10, 22, 39, 54, 65 |
| ei 5 | 13, 16, 42, 57, 67 |

Table A-2. Parcel items for PM (practical-imaginative).

| Parcel | Items in Parcel |
|--------|-----------------|
| pm 1 | 3, 6, 9, 64 |
| pm 2 | 11, 21, 31, 41 |
| pm 3 | 15, 29, 34, 48 |
| pm 4 | 25, 45, 51, 68 |

Table A-3. Parcel items for TF (thinking-feeling).

| Parcel | Items in Parcel |
|--------|--------------------|
| tf 1 | 12, 18, 29, 37, 55 |
| tf 2 | 24, 28, 43, 50, 61 |

Table A-4. Parcel items for OL (organized-flexible).

| Parcel | Items in Parcel |
|--------|--------------------|
| ol1 | 2, 23, 32, 53 |
| ol 2 | 5, 38, 39, 56 |
| ol3 | 8, 10, 40, 41, 58 |
| ol 4 | 14, 27, 47, 59, 63 |
| ol 5 | 17, 30, 35, 66 |
| ol6 | 20, 44, 54, 69 |

APPENDIX B
STANDARDIZED REGRESSION WEIGHTS AND LATENT FACTOR
CORRELATIONS FOR THE PARCEL MODEL

Table B-1. Standardized regression weights for the Australian sample.

| Parcel | EI | OL | PM | TF |
|--------|------|------|-----|------|
| ei1 | .02 | | | |
| ei2 | .10 | | | |
| ei3 | -.36 | | | |
| ei4 | -.17 | | | |
| ei5 | -.12 | | | |
| ol1 | | .34 | | |
| ol2 | | .70 | | |
| ol3 | | .08 | | |
| ol4 | | .19 | | |
| ol5 | | -.55 | | |
| ol6 | | .66 | | |
| pm1 | | | .19 | |
| pm2 | | | .33 | |
| pm3 | | | .47 | |
| pm4 | | | .12 | |
| tf1 | | | | 1.00 |
| tf2 | | | | -.02 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-2. Standardized regression weights for the Chinese sample.

| Parcel | EI | OL | PM | TF |
|--------|------|------|-----|-----|
| ei1 | .00 | | | |
| ei2 | -.03 | | | |
| ei3 | -.25 | | | |
| ei4 | -.18 | | | |
| ei5 | -.00 | | | |
| ol1 | | .35 | | |
| ol2 | | .82 | | |
| ol3 | | .17 | | |
| ol4 | | .24 | | |
| ol5 | | -.47 | | |
| ol6 | | .53 | | |
| pm1 | | | .15 | |
| pm2 | | | .34 | |
| pm3 | | | .07 | |
| pm4 | | | .32 | |
| tf1 | | | | .06 |
| tf2 | | | | .37 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-3. Standardized regression weights for the Costa Rican sample.

| Parcel | EI | OL | PM | TF |
|--------|------|------|-----|-----|
| ei1 | .03 | | | |
| ei2 | .11 | | | |
| ei3 | -.17 | | | |
| ei4 | -.12 | | | |
| ei5 | -.06 | | | |
| ol1 | | .47 | | |
| ol2 | | .62 | | |
| ol3 | | -.06 | | |
| ol4 | | -.08 | | |
| ol5 | | -.70 | | |
| ol6 | | .76 | | |
| pm1 | | | .24 | |
| pm2 | | | .24 | |
| pm3 | | | .69 | |
| pm4 | | | .08 | |
| tf1 | | | | .35 |
| tf2 | | | | .09 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-4. Standardized regression weights for the Gazan sample.

| Parcel | EI | OL | PM | TF |
|--------|------|------|------|------|
| ei1 | .00 | | | |
| ei2 | -.02 | | | |
| ei3 | -.04 | | | |
| ei4 | .09 | | | |
| ei5 | 1.00 | | | |
| ol1 | | .32 | | |
| ol2 | | .45 | | |
| ol3 | | .21 | | |
| ol4 | | -.20 | | |
| ol5 | | -.57 | | |
| ol6 | | .69 | | |
| pm1 | | | .54 | |
| pm2 | | | .52 | |
| pm3 | | | -.09 | |
| pm4 | | | -.11 | |
| tf1 | | | | 1.00 |
| tf2 | | | | .07 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-5. Standardized regression weights for the Nigerian sample.

| Parcel | EI | OL | PM | TF |
|--------|------|------|------|-----|
| ei1 | .01 | | | |
| ei2 | .05 | | | |
| ei3 | 1.00 | | | |
| ei4 | .02 | | | |
| ei5 | -.15 | | | |
| ol1 | | .35 | | |
| ol2 | | .41 | | |
| ol3 | | .08 | | |
| ol4 | | .04 | | |
| ol5 | | -.52 | | |
| ol6 | | .64 | | |
| pm1 | | | 1.00 | |
| pm2 | | | .12 | |
| pm3 | | | -.01 | |
| pm4 | | | .06 | |
| tf1 | | | | .14 |
| tf2 | | | | .45 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-6. Standardized regression weights for the Philippine sample.

| Parcel | EI | OL | PM | TF |
|--------|------|------|------|-----|
| ei1 | .00 | | | |
| ei2 | .00 | | | |
| ei3 | 1.00 | | | |
| ei4 | -.01 | | | |
| ei5 | -.02 | | | |
| ol1 | | .36 | | |
| ol2 | | .67 | | |
| ol3 | | .20 | | |
| ol4 | | -.18 | | |
| ol5 | | -.59 | | |
| ol6 | | .33 | | |
| pm1 | | | 1.00 | |
| pm2 | | | .17 | |
| pm3 | | | -.04 | |
| pm4 | | | .09 | |
| tf1 | | | | .08 |
| tf2 | | | | .27 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-7. Standardized regression weights for the USA sample.

| Parcel | EI | OL | PM | TF |
|--------|------|------|------|------|
| ei1 | .00 | | | |
| ei2 | .03 | | | |
| ei3 | -.26 | | | |
| ei4 | -.06 | | | |
| ei5 | -.04 | | | |
| ol1 | | .04 | | |
| ol2 | | .55 | | |
| ol3 | | -.04 | | |
| ol4 | | -.32 | | |
| ol5 | | -.60 | | |
| ol6 | | .64 | | |
| pm1 | | | .01 | |
| pm2 | | | .03 | |
| pm3 | | | 3.40 | |
| pm4 | | | -.01 | |
| tf1 | | | | .05 |
| tf2 | | | | -.25 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-8. Standardized regression weights for the Zimbabwe sample.

| Parcel | EI | OL | PM | TF |
|--------|------|------|------|------|
| ei1 | .09 | | | |
| ei2 | -.08 | | | |
| ei3 | -.24 | | | |
| ei4 | -.30 | | | |
| ei5 | .02 | | | |
| ol1 | | .13 | | |
| ol2 | | .38 | | |
| ol3 | | .24 | | |
| ol4 | | .09 | | |
| ol5 | | -.37 | | |
| ol6 | | .55 | | |
| pm1 | | | 1.97 | |
| pm2 | | | .09 | |
| pm3 | | | .02 | |
| pm4 | | | .09 | |
| tf1 | | | | .05 |
| tf2 | | | | -.75 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-9. Correlations among the four bipolar styles for the Australian sample.

| | EI | OL | PM | TF |
|----|----|-------|------|------|
| EI | | -1.12 | -.33 | .07 |
| OL | | | .08 | -.06 |
| PM | | | | .42 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-10. Correlations among the four bipolar styles for the Chinese sample.

| | EI | OL | PM | TF |
|----|----|-------|------|-------|
| EI | | -1.97 | -.33 | -1.34 |
| OL | | | -.40 | -.79 |
| PM | | | | 1.21 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-11. Correlations among the four bipolar styles for the Costa Rican sample.

| | EI | OL | PM | TF |
|----|----|-------|------|------|
| EI | | -2.09 | -.36 | .21 |
| OL | | | -.15 | -.16 |
| PM | | | | 1.30 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-12. Correlations among the four bipolar styles for the Gaza sample.

| | EI | OL | PM | TF |
|----|----|------|------|-----|
| EI | | -.43 | -.14 | .07 |
| OL | | | .72 | .02 |
| PM | | | | .19 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-13. Correlations among the four bipolar styles for the Nigerian sample.

| | EI | OL | PM | TF |
|----|----|-----|------|-----|
| EI | | .27 | -.01 | .56 |
| OL | | | -.08 | .13 |
| PM | | | | .02 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-14. Correlations among the four bipolar styles for the Philippine sample.

| | EI | OL | PM | TF |
|----|----|-----|-----|------|
| EI | | .33 | .04 | .82 |
| OL | | | .09 | -.88 |
| PM | | | | -.19 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-15. Correlations among the four bipolar styles for the USA sample.

| | EI | OL | PM | TF |
|----|----|-------|------|------|
| EI | | -1.74 | .05 | 3.16 |
| OL | | | -.06 | -.18 |
| PM | | | | .06 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table B-16. Correlations among the four bipolar styles for the Zimbabwe sample.

| | EI | OL | PM | TF |
|----|----|-------|------|-----|
| EI | | -1.68 | .05 | .69 |
| OL | | | -.02 | .05 |
| PM | | | | .02 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

APPENDIX C
CONTENT OF MODIFIED PARCELS

Sociability

- 4. After school I like to do something (alone, with friends)
- 57. When I have free time, I like to (do something with friends, spend time by myself)
- 62. When I get home from school (I want to talk to someone about my day, I want to be alone for a while)
- 67. I most like to (read or watch TV, talk with friends)

Restraint

- 24. Most people are likely to say I am (brave, shy)
- 30. When I see new students in my class (I introduce myself to them, I stand back and watch them for a while)
- 49. I make new friends (more quickly than others, more slowly than others)
- 52. If others are playing a new game, (I watch them for a while, I jump right in and join them)

Assertiveness

- 1. When in a large group, I most often (talk, listen)
- 26. I often (tell people what I really feel, keep my feelings to myself)
- 46. I most often (keep my real thoughts to myself, tell people what I am really thinking)
- 54. In class, I talk (a lot, a little)

Introspection

- 10. In school I prefer (quiet seatwork, active work in groups)
- 19. If I need to think about something important, (I find a friend I can talk with, I go for a walk alone)
- 36. I like sports that (I can do by myself like biking and swimming, I can play with a team like volleyball and baseball)
- 42. In school I like to work (in a group, on my own)

Openness

- 13. I like to have (a few good friends, lots of friends)
- 16. I almost always (hide my feelings, show my feelings)
- 33. When something really good happens to me, (I am pleased but keep calm, I get very excited and show it)
- 39. When planning my birthday party, I want to invite (a few good friends, many friends)
- 65. When I'm with friends (I often listen to others who tell jokes, I'm often the one who tells jokes)

Discipline

- 5. When doing school work, usually (I start it just before it's due, I finish it long before it's due)
- 20. When my parents or teachers give me many jobs, (I start with the one I like and don't plan the order, I make a list of them to do in order)
- 23. Spelling or math contests (are fun and make students want to learn, are mean because the losers feel bad)
- 27. It is more important (to like what I am doing, to finish what I start)
- 38. If I have work to do, I like to do it (after I play, before I play)
- 60. I do my homework (at the last minute, as soon as possible)

Flexibility

- 8. I like to work on (one thing at a time, many things at the same time)
- 14. When working, I most often want (to do a good job, to please others)
- 35. I want my work to be (done correctly, pleasing to others)
- 47. When practicing something like a musical instrument or a sport, (I carefully watch the time to be sure I practice enough, I lose track of the time and have fun with what I am doing)
- 56. When walking to a friend's house, (I sometimes wonder around on my way there, I walk straight there)
- 58. The way I see it, people should (make their work into play, work before they play)
- 63. When do something on a computer, (I just use it, I wonder how it works)

Organization

- 2. When planning to be away from home for a few days, (I put stuff in my suitcase right before I leave, I make lists of what I want to take)
- 17. I like my desk to be (clean and orderly, any old way)
- 32. After playing with games, (I leave them out, I put them back where they belong)
- 44. When friends spend the night at my house, (I just see what happens when we are together, I plan what we will do together)
- 66. In my room (everything has its place, I don't care if I have to search for things)
- 69. When I have school work, I most often (need to be reminded when it's due, remember when it's due)

Conscientiousness

- 35. I want my work to be (done correctly, pleasing to others)
- 40. When I spend my money on something, (I think about it for a long time, I decide quickly)
- 53. When going to someone's home, I like to be there (a little before or after they said, at the exact time they said)
- 59. If I'm in trouble with my parents, (I try to talk my way out of it, I keep quiet)

Creativity

- 3. I do my chores at home (the way most people do them, in my own way)
- 11. My ideas (are about the same as those of others, are often different than those of others)
- 31. I do my homework (the way most students do it, in my own way)
- 45. I look up to people most who (have ideas that are different, have ideas that work)
- 48. I like to be with people who have (practical ideas, unusual ideas)
- 64. I like to play games that (already have rules, let us make up rules as we go along)

Liberalism

- 6. In school I like to learn about (ideas that make me think in new ways, facts that help me know lots of things)
- 21. I like TV programs that (teach me different ways, teach me more facts)
- 29. Teachers are best who (teach me to think in new ways, give me lots of information)
- 68. My head seems filled with (big ideas, lots of little facts)

Ingenuity

- 9. I enjoy listening to others tell stories that are (made up, true)
- 15. I like school subjects that (give me usable skills, spark my imagination)
- 25. I like to write about something that (really happened, I make up)
- 34. I prefer stories that (are full of detail, let my imagination wonder)
- 41. For me, using my imagination (is boring, is fun)
- 51. I like to hear stories that (others make up, really happened to others)

Sensitivity

- 12. When people disappoint me, (I forget it soon, I take a while to get over my hurt feelings)
- 18. When it comes to the sad part of movies, (I feel sad, I'm not really bothered)
- 28. When people say mean things to me, even when they are teasing, (they hurt my feelings, they do not bother me much)
- 61. If someone breaks a promise, (I feel hurt for a long time, I get over it quickly)

Passion

- 37. When I see someone who is upset, I (get upset too, believe they'll get over it)
- 55. When I'm in a group that is upset because of some bad news, (I cry right with them, I don't cry because someone has to be strong)
- 43. If I could, I would prefer to help (design a house, build a house)
- 50. I decide on things based on what's in my (head, heart)

APPENDIX D
STANDARDIZED REGRESSION WEIGHTS AND LATENT FACTOR
CORRELATIONS FOR THE MODIFIED MODEL

Table D-1. Standardized regression weights for the Australian sample.

| Quality | EI | OL | PM | TF |
|-------------------|------|-----|------|-----|
| Sociability | .16 | | | |
| Restraint | .07 | | | |
| Assertiveness | -.09 | | | |
| Introspection | .22 | | | |
| Openness | .03 | | | |
| Sensitivity | | | | .88 |
| Passion | | | | .22 |
| Discipline | | .87 | | |
| Flexibility | | .11 | | |
| Organization | | .26 | | |
| Conscientiousness | | .14 | | |
| Creativity | | | .24 | |
| Liberalism | | | -.69 | |
| Ingenuity | | | .26 | |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table D-2. Standardized regression weights for the Chinese sample.

| Quality | EI | OL | PM | TF |
|-------------------|------|-----|------|-----|
| Sociability | .18 | | | |
| Restraint | .55 | | | |
| Assertiveness | .35 | | | |
| Introspection | .07 | | | |
| Openness | -.39 | | | |
| Sensitivity | | | | .39 |
| Passion | | | | .51 |
| Discipline | | .26 | | |
| Flexibility | | .74 | | |
| Organization | | .21 | | |
| Conscientiousness | | .37 | | |
| Creativity | | | .39 | |
| Liberalism | | | -.18 | |
| Ingenuity | | | .56 | |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table D-3. Standardized regression weights for the Costa Rican sample.

| Quality | EI | OL | PM | TF |
|-------------------|------|-----|------|-----|
| Sociability | .19 | | | |
| Restraint | .20 | | | |
| Assertiveness | -.20 | | | |
| Introspection | .64 | | | |
| Openness | -.14 | | | |
| Sensitivity | | | | .35 |
| Passion | | | | .35 |
| Discipline | | .66 | | |
| Flexibility | | .19 | | |
| Organization | | .27 | | |
| Conscientiousness | | .27 | | |
| Creativity | | | 1.32 | |
| Liberalism | | | .05 | |
| Ingenuity | | | .08 | |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table D-4. Standardized regression weights for the Gazan sample.

| Quality | EI | OL | PM | TF |
|-------------------|------|-----|-----|-----|
| Sociability | .06 | | | |
| Restraint | -.22 | | | |
| Assertiveness | -.60 | | | |
| Introspection | -.10 | | | |
| Openness | .35 | | | |
| Sensitivity | | | | .31 |
| Passion | | | | .46 |
| Discipline | | .54 | | |
| Flexibility | | .25 | | |
| Organization | | .35 | | |
| Conscientiousness | | .22 | | |
| Creativity | | | .23 | |
| Liberalism | | | .25 | |
| Ingenuity | | | .24 | |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table D-5. Standardized regression weights for the Nigerian sample.

| Quality | EI | OL | PM | TF |
|-------------------|------|-----|------|-----|
| Sociability | .05 | | | |
| Restraint | 1.27 | | | |
| Assertiveness | .02 | | | |
| Introspection | .10 | | | |
| Openness | -.21 | | | |
| Sensitivity | | | | .26 |
| Passion | | | | .45 |
| Discipline | | .10 | | |
| Flexibility | | .44 | | |
| Organization | | .11 | | |
| Conscientiousness | | .38 | | |
| Creativity | | | .41 | |
| Liberalism | | | -.21 | |
| Ingenuity | | | .21 | |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
 TF = Thinking-Feeling.

Table D-6. Standardized regression weights for the Philippine sample.

| Quality | EI | OL | PM | TF |
|-------------------|------|------|------|-----|
| Sociability | .11 | | | |
| Restraint | .67 | | | |
| Assertiveness | .04 | | | |
| Introspection | .09 | | | |
| Openness | -.39 | | | |
| Sensitivity | | | | .31 |
| Passion | | | | .09 |
| Discipline | | 1.50 | | |
| Flexibility | | .00 | | |
| Organization | | .13 | | |
| Conscientiousness | | -.09 | | |
| Creativity | | | .16 | |
| Liberalism | | | -.02 | |
| Ingenuity | | | -.18 | |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
 TF = Thinking-Feeling.

Table D-7. Standardized regression weights for the USA sample.

| Quality | EI | OL | PM | TF |
|-------------------|------|------|------|-----|
| Sociability | .04 | | | |
| Restraint | -.22 | | | |
| Assertiveness | -.55 | | | |
| Introspection | .02 | | | |
| Openness | .32 | | | |
| Sensitivity | | | | .13 |
| Passion | | | | .62 |
| Discipline | | .31 | | |
| Flexibility | | -.35 | | |
| Organization | | -.06 | | |
| Conscientiousness | | -.54 | | |
| Creativity | | | .17 | |
| Liberalism | | | -.49 | |
| Ingenuity | | | .41 | |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table D-8. Standardized regression weights for the Zimbabwe sample.

| Quality | EI | OL | PM | TF |
|-------------------|------|------|------|------|
| Sociability | .31 | | | |
| Restraint | .50 | | | |
| Assertiveness | -.06 | | | |
| Introspection | .03 | | | |
| Openness | -.06 | | | |
| Sensitivity | | | | .27 |
| Passion | | | | -.13 |
| Discipline | | 1.00 | | |
| Flexibility | | -.02 | | |
| Organization | | .18 | | |
| Conscientiousness | | -.03 | | |
| Creativity | | | 1.00 | |
| Liberalism | | | .07 | |
| Ingenuity | | | .07 | |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table D-9. Correlations among the four bipolar styles for the Australian sample.

| | EI | OL | PM | TF |
|----|----|-----|------|------|
| EI | | .00 | .00 | .91 |
| OL | | | -.47 | -.10 |
| PM | | | | .15 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table D-10. Correlations among the four bipolar styles for the Chinese sample.

| | EI | OL | PM | TF |
|----|----|------|------|-----|
| EI | | -.15 | -.27 | .43 |
| OL | | | .12 | .13 |
| PM | | | | .28 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table D-11. Correlations among the four bipolar styles for the Costa Rican sample.

| | EI | OL | PM | TF |
|----|----|------|-----|------|
| EI | | -.52 | .15 | .33 |
| OL | | | .11 | -.19 |
| PM | | | | -.03 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table D-12. Correlations among the four bipolar styles for the Gaza sample.

| | EI | OL | PM | TF |
|----|----|------|------|------|
| EI | | -.63 | -.56 | .23 |
| OL | | | 1.06 | -.12 |
| PM | | | | .64 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table D-13. Correlations among the four bipolar styles for the Nigerian sample.

| | EI | OL | PM | TF |
|----|----|------|-----|------|
| EI | | -.15 | .17 | .13 |
| OL | | | .24 | -.12 |
| PM | | | | .44 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table D-14. Correlations among the four bipolar styles for the Philippine sample.

| | EI | OL | PM | TF |
|----|----|------|-----|------|
| EI | | -.20 | .02 | .13 |
| OL | | | .37 | -.12 |
| PM | | | | .44 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table D-15. Correlations among the four bipolar styles for the USA sample.

| | EI | OL | PM | TF |
|----|----|------|------|------|
| EI | | -.36 | .19 | -.10 |
| OL | | | -.13 | .12 |
| PM | | | | -.36 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

Table D-16. Correlations among the four bipolar styles for the Zimbabwe sample.

| | EI | OL | PM | TF |
|----|----|------|------|------|
| EI | | -.09 | .32 | .77 |
| OL | | | -.05 | -.31 |
| PM | | | | .29 |

EI = Extroverted-Introverted. OL = Organized-Flexible. PM = Practical-Imaginative.
TF = Thinking-Feeling.

LIST OF REFERENCES

- Agresti, A. & Finlay, B. (1997). *Statistical Methods for the Social Sciences (3rd edition)*. Upper Saddle River, NJ: Prentice Hall.
- Aldenerfer, M.S., & Blashfield, R.K. (1984). *Cluster analysis*. (Sage University Paper series on Quantitative Applications in the Social Sciences, No. 07-044). Newbury Park, CA.
- Allport, G.W., & Odbert, H.S. (1936). Trait names: A psycholexical study. *Psychological Monographs*, 47, 63.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.
- Anastasi, A. & Urbina, S. (1997). *Psychological Testing. (7th Edition)*. Upper Saddle River, NJ: Prentice Hall.
- Angleitner, A., & Ostendorf, F. (1994). Temperament and the Big Five Factors of personality. In C.F. Halverson, G.A. Kohstamm, and R.P. Martin (Eds.), *The developing structures of temperament and personality from infancy to adulthood*. Hillsdale, NJ: Erlbaum.
- Arbuckle, J.L. (2003). *Amos 5.0* [Computer software]. Chicago: Smallwaters.
- Aune, K.S., & Aune, R.K. Cultural differences in the self-reported experience and expression of emotions in relationships. *Journal of Cross-Cultural Psychology*, 27, 67-81.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bassett, K.D. (2001). *Nature, nurture, and temperament: Cross-national comparisons of temperament styles displayed by Australian, Chinese, and United States students*. Unpublished master's thesis. University of Florida, Gainesville.
- Bassett, K.D. (2004). *Nature, nurture, and temperament: Comparisons of temperament styles displayed by United States students*. Unpublished doctoral dissertation. University of Florida, Gainesville.

- Bates, J.E., & Wachs, T.D. (1994). Preface. In J.E. Bates and T.D. Wachs (Eds.), *Temperament: Individual differences at the interface of biology and behavior*. Washington, D.C.: American Psychological Association.
- Bates, J.E., & Wachs, T.D., & Emde (1994). Toward practical uses for biological concepts of temperament. In J.E. Bates and T.D. Wachs (Eds.), *Temperament: Individual differences at the interface of biology and behavior*. Washington, D.C.: American Psychological Association.
- Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practices*, 17, 10-17.
- Bess, T.L., Harvey, R.J., & Swartz, D. (2003). *Hierarchical confirmatory factor analysis of the Myers-Briggs Type Indicator*. Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Bouchard, T.J. (1996). Genetics and evolution: Implications for personality theories. In J. Newman (Ed.), *Measures of the Five Factor Model and psychological type: A major convergence of research and theory*. (pp. 19-40). Gainesville, FL: Center for Applications of Psychological Type.
- Broer, E., & McCarley, N. (1999). Using and validating the Myers-Briggs Type Indicator in mainland China. *Journal of Psychological Type*, 51, 5-21.
- Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.) *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Buss, A.H., & Plomin, R. (1975). A temperament theory of personality development. In K. MacCorquodale (Series Ed.), *Wiley Series in Behavior*. New York: Wiley.
- Buss, A.H., & Plomin, R. (1984). *Temperament: Early developing personality traits*. Hillsdale, NJ: Erlbaum.
- Byrne, B.M. (1998). *Structural Equation Modeling with Lisrel, Prelis, and Simplis: Basic Concepts, Applications, and Programming*. Mahwah, NJ: Erlbaum.
- Byrne, B.M. (2001). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. Mahwah, NJ: Erlbaum.
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminate validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Campos, J.J., Barrett, K.C., Lamb, M.E., Goldsmith, H.H., & Stenberg, C. (1983). Socioemotional development. In M.M. Haith & J.J. Campos (Eds.), P.H. Mussen (Series Ed.), *Handbook of child psychology: Vol. 2. Infancy and developmental psychobiology* (pp. 783-915). New York: Wiley.

- Carey, W.B. (1970). A simplified method for measuring infant temperament. *Journal of Pediatrics*, 77, 188-194.
- Carey, W.B., & McDevitt, S.C. (1978). *Revision of the Infant Temperament Questionnaire*. *Pediatrics*, 61, 735-739.
- Carlson, S.R. (2001). *Cross-cultural differences in the temperament patterns of children from China, Costa Rica, and the United States*. Unpublished Masters Thesis. University of Florida, Gainesville.
- Cattell, R.B. (1982). *The Inheritance of Personality and Ability: Research Methods and Findings*. New York: Academic Press.
- Cattell, R.B., & Gibbons, B.D. (1968). Personality factor structure of the combined Guilford and Cattell personality questionnaires. *Journal of Personality and Social Psychology*, 9, 107-120.
- Cattell, R.B., & Kline, P. (1977). The scientific analysis of personality and motivation. In D.T. Lykken (Series Ed.), *Personality and psychopathology: A series of monographs, texts, and treatises*. New York: Academic Press.
- Cheung, F.M., Leung, K. Fan, R.M., Son, W.Z., Zhang, J.X., & Zhang, J.P. (1996). Development of the Chinese Personality Assessment Inventory. *Journal of Cross-Cultural Psychology*, 27, 181-199.
- Church, A.T. (2000). Culture and personality: Toward an integrated cultural trait psychology. *Journal of Personality*, 69, 651-703.
- Church, A.T., Katigbak, M.S. & Reyes, J.A. (1998). Further exploration of Filipino personality structure using the lexical approach: Do the Big-Five or Big Seven dimensions emerge? *European Journal of Personality*, 12, 249-270.
- Church, A.T., & Lonner, W.J. (1998). The cross-cultural perspective in the study of personality: rationale and current research. *Journal of Cross-Cultural Psychology*, 29, 32-62.
- Costa, P.T. & McCrae, R.R. (1985). *NEO-PI Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Harcourt Brace Jovanovich.
- Cronbach, L.J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.
- Cronbach, L.J., & Gleser, G.C. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana: University of Illinois Press.

- deVries, M.W. (1984). Temperament and infant mortality among the Masai of East Africa. *American Journal of Psychiatry*, *141*, 1189-1194.
- Diener, E. (2000). Introduction to the special section of personality differences. *Journal of Personality and Social Psychology*, *78*, 120-121.
- Eysenck, H.J. (1982). Personality, genetics, and behavior: Selected papers. In C.D. Spielberger, (General Ed.), *Centennial Psychology Series*. New York: Praegar.
- Eysenck, H.J., & Eysenck, M.W. (1985). *Personality and individual differences: A natural science approach*. New York: Plenum.
- Fijneman, Y.A., Willemsen, M.E., Poortinga, Y.H. (1996). Individualism-collectivism: An empirical study of a conceptual issue. *Journal of Cross-Cultural Psychology*, *27*, 381-402.
- Garrison, W.T., Earls, F.J. (1987). Temperament and Child Psychopathology. In A. E. Kazdin (Series Ed.), *Developmental Clinical Psychology and Psychiatry Series: Vol. 2*.
- Glutting, J.J., Youngstrom, E.A., Oakland, T., Watkins, M.W. (1996). Situational specificity and generality of test behaviors for samples of normal and referred children. *School Psychology Review*, *25*, 94-107.
- Goldsmith, H.H., Buss, A.H., Plomin, R., Rothbart, M.K., Thomas, A., Chess, S., Hinde, R.A., & McCall, R.B. (1987). Roundtable: What is temperament? Four approaches. *Child Development*, *58*, 505-529.
- Goldsmith, H.H., Lemery, K.S., Aksan, N., & Buss, K.A. (2000). Temperamental substrates of personality development. In V.I. Molfese and D.L. Molfese (Eds.), *Temperament and personality development across the life span* (pp. 1-32). Mahway, NJ: Lawrence Erlbaum Associates.
- Goldsmith, H.H., & Rieser-Danner, L. (1986). Variation among temperament theories and validation studies of temperament assessment. In G.A. Kohnstamm (Ed.), *Temperament discussed: Temperament and development in infancy and childhood* (pp. 1-9). Berwyn, PA: Swets North America.
- Gosling, S.D., & John, O.P. (1998). Personality dimensions in dogs, cats, and hyenas. In S.D. Gosling & S. Suomi (Chairs). *From mice to men: Bridging the gap between personality and animal researchers*. Symposium conducted at the 10th Annual Convention of the American Psychological Society. Washington, DC.
- Gray, J.A. (1991). The neuropsychology of temperament. In J. Stelau & A. Angleitner (Eds.), *Explorations in temperament: International perspectives on theory and measurement* (pp. 105-128). New York: Plenum.

- Guilford, J.S., Zimmerman, W.S., & Guilford, J.P. (1976) *The Guilford-Zimmerman Temperament Survey Handbook*. San Diego: Edits.
- Gunnar, M.R. (1990). The psychobiology of infant temperament. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction* (pp. 387-410). Hillsdale, NJ: Erlbaum.
- Hambleton, R.K. (2003). Issues, designs, and technical guidelines for adapting tests in multiple languages and cultures. In R.K. Hambleton, P. Merenda, & C. Spielberger (Eds.), *Adapting educational and psychological testing for cross-cultural assessment*. Hillsdale, NJ: Lawrence Erlbaum Publishers.
- Hambleton, R.K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. As seen in the *Journal of Applied Testing Technology*, 1, 1-30. Retrieved February 28, 2005, from <http://www.testpublishers.org/Documents/journal0114.pdf>
- Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Dordrecht: Kluwer-Nijhoff.
- Henderson, L.J. (1913). *The fitness of the environment*. New York: Macmillan.
- Holland, P.W., & Thayer, D.T. (1988). Differential item performance and the Matel-Haenszel procedure. In H. Wainer & H.I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.
- Horton, C.B., Oakland, T. (1997). Temperament-based learning styles as moderators of academic achievement. *Adolescence*, 32, 131-141.
- Hsu, F. (1985). The self in cross-cultural perspective. In A.J. Marsella, G. DeVos, & F. Hsu (Eds.), *Culture and self: Asian and Western perspectives*. New York: Tavistock Press.
- International Conference on Test Adaptation: Adapting tests for use in multiple languages and cultures: May 20-22, 1999. Georgetown University Conference Center, Washington D.C.
- Joreskog, K.G. (1969). A general approach to confirmatory factor analysis. *Psychometrika*, 34, 183-202.
- Jung, C.G. (1971). *Psychological types*. (R.F.C. Hull, Revision of Trans. By H.G. Baynes). Princeton, NJ: Princeton University Press. (Original work published 1921).
- Kagan, J. (1989). Temperamental contributions to social behavior. *American Psychologist*, 44, 668-674.

- Kagan, J. (1994). *Galen's Prophecy: Temperament in human nature*. New York: Basic Books.
- Keirse, D. (1998). *Please understand me II: Temperament, character, intelligence*. Del Mar, CA: Prometheus Nemesis.
- Keith, T.Z. (1997). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D.P. Flanagan, J.L. Genshaft, & P.L. Harrison. (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 373-402). New York: Guilford.
- Keith, T.Z., & Kranzler, J.H. (1999). The absence of structural fidelity precludes construct validity: Rejoinder to Naglieri on what the Cognitive Assessment System does and does not measure. *School Psychology Review*, 28, 303-321.
- Keogh, B.K. (2003). *Temperament in the classroom: Understanding individual differences*. Baltimore, MD: Brookes.
- King, J.E., & Figueredo, A.J. (1997). The five-factor model plus dominance in chimpanzee personality. *Journal of Research in Personality*, 31, 257-271.
- Kline, R.B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.
- Kluckhohn, C. (1954). Culture and behavior. In G. Lidzey (Ed.), *Handbook of Social Psychology* (2nd ed., pp. 921-976). Cambridge, MA: Addison-Wesley.
- Kluckhohn, C., & Murray, H.A. (1953). Personality formation: The determinants. In C. Kluckhohn, H.A. Murray, & D.M. Schneider (Eds.), *Personality in nature, society, and culture* (2nd ed., pp. 53-67). New York: Knopf.
- Kohnstamm, G.A. (1989). Temperament in childhood: Cross-cultural and sex differences. In G.A. Kohnstamm, J.E. Bates, & M.K. Rothbart (Eds.), *Temperament in childhood* (pp. 483-508). Chichester, England: Wiley.
- Kruskal, J.B., & Wish, M. (1978). *Multidimensional scaling*. (Sage University Paper series on Quantitative Applications in the Social Sciences, No. 07-011).
- Lemery, K.S., Goldsmith, H.H., Klinner, M.D., & Mrazek, D.A. (1999). Developmental models of infant and childhood temperament. *Developmental Psychology*, 35, 189-204.
- Lombardo, G.P., & Foschi, R. (2002). The European origins of "personality psychology". *European Psychologist*, 7, 134-145.

- Long, J.S., (1983a). *Confirmatory factor analysis: A preface to Lisrel*. (Sage University Paper series on Quantitative Applications in the Social Sciences, No. 07-033). Newbury, CA.
- Long, J.S., (1983b). *Covariance structure models: An introduction to Lisrel*. (Sage University Paper series on Quantitative Applications in the Social Sciences, No. 07-034) Newbury, CA.
- Macdaid, G.P., McCaulley, M.H., & Kainz, R.I. (1986). *Atlas of type tables*. Gainesville, FL: Center for Applications of Psychological Type.
- Macoby, E.E., & Jacklin, C.N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press
- Marsella, A.J., & Leong, F.T. (1995). Cross-cultural issues in personality and career assessment. *Journal of Career Assessment*, 3, 202-218.
- Martin, R. (1988). *Assessment of personality and behavior problems: Infancy through adolescence*. New York: Guilford Press.
- Martin, R.P. Wisenbaker, J., & Huttunen, M. (1994). Review of factor analytic studies of temperament measures based on the Thomas-Chess structural model: Implications for the big Five. In C.F. Halverson, G.A. Kohnstamm, and R.P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 157-172).
- McCaulley, M.H., & Moody, R.A. (2001). Multicultural applications of the Myers-Briggs Type Indicator. In L.A. Suzuki, J.G. Ponterotto, & P.J. Meller (Eds.), *Handbook of Multicultural Assessment: Clinical, Psychological, and Educational Applications (2nd ed.)*. (pp. 279-306). San Francisco: Jossey-Bass.
- McCrae, R.R., & Costa, P.T., (1997). Personality trait structure as a human universal. *American Psychologist*, 52, 509-516.
- McCrae, R.R., & Costa, P.T., (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the Five Factor Model of personality. *Journal of Personality*, 57, 17-40.
- McCrae, R.R., Costa, P.T., Ostendorf, F. Angleitner, A., Hrebickova, M., Avia, M.D., Sanz, J., & Sanchez-Bernardos, M.L. (2000). Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology*, 78, 173-186.
- Merenda, P.F. (2004). Cross-cultural adaptation of educational and psychological testing. In Hambleton, R., Meranda, P., & Spielberger, C. (Eds.). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NY: Erlbaum.

- Merenda, P.F. (1997). Methods, plainly speaking: A guide to the proper use of factor analysis in the conduct and reporting of research: Pitfalls to avoid. *Measurement and Evaluation in Counseling and Development*, 30, 156-164.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York: Macmillan.
- Millon, T. (1991). Classification in psychopathology: Rationale, alternatives, and standards. *Journal of Abnormal Psychology*, 100, 245-261.
- Myers, I.B. & McCaulley, M.H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. (2nd ed.). Palo Alto, CA: Davies-Black.
- Myers, I.B., McCaulley, M.H., Quenk, N.L., & Hammer, A.L. (1998). *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Nelson, C.A. (1994). Neural bases of Infant development. In J.E. Bates and T.D. Wachs (Eds.), *Temperament: Individual differences at the interface of biology and behavior*. (pp. 47-82). Washington, D.C.: American Psychological Association.
- Newman, J. (1996a). A brief history of the science of personality. In J. Newman (Ed.), *Measures of the Five Factor Model and psychological type: A major convergence of research and theory*. (pp. 1-18). Gainesville, FL: Center for Applications of Psychological Type.
- Newman, J. (1996b). Trait theory, type theory and the biological basis of personality. In J. Newman (Ed.), *Measures of the Five Factor Model and psychological type: A major convergence of research and theory*. (pp. 63-80). Gainesville, FL: Center for Applications of Psychological Type.
- Norman, W.T. (1963). Towards an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 66, 574-583.
- Oakland, T. (2002). *Temperament styles of students in five countries*. Paper presented at the annual meeting of the International Association of Applied Psychology, Singapore.
- Oakland, T. (2004). Selected ethical issues relevant to test adaptations. In Hambleton, R., Merenda, P., & Spielberger, C. (Eds.). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NY: Erlbaum.
- Oakland, T., Glutting, J.J., & Horton, C.B. (1996). *Student Styles Questionnaire Manual*. San Antonio TX: Psychological Corporation.

- Oakland, T., Stafford, M.E., Horton, C.B., & Glutting, J.J. (2002). Temperament and vocational preferences: Age, gender, and racial-ethnic comparisons using the Student Styles Questionnaire. *Journal of Career Assessment, 9*, 97-314.
- Osterlin, S.J., Miao, D., Sheng, Y., & Chia, R. (2004). Adapting item format for cultural effects in translated tests: Cultural effects on construct validity of the Chinese versions of the MBTI. *International Journal of Testing, 4*, 61-74.
- Pap, A. (1953). Reduction-sentences and open concepts. *Methodos, 5*, 3-30.
- Piker, S. (1998). Contributions to psychological anthropology. *Journal of Cross-Cultural Psychology, 29*, 9-31.
- Pittenger, D.J. (1993). The utility of the Myers-Briggs Type Indicator. *Review of Educational Research, 63*, 467-488.
- Plomin, R., & Saudino, K.J. (1994). Quantitative genetics and molecular genetics. In J.E. Bates and T.D. Wachs (Eds.), *Temperament: Individual differences at the interface of biology and behavior*. (pp. 143-173). Washington, D.C.: American Psychological Association.
- Quenk, N. (1993). Personality types or personality traits: what difference does it make? *Bulletin of Psychological Type, 16*, 11-14.
- Quenk, N. (2000). Essentials of Myers-Briggs Type Indicator Assessment. In A.S. Kaufman and N.L. Kaufman (Series Eds.), *Essentials of Psychological Assessment*. New York: Wiley.
- Riemann, R., Angleitner, A., & Strelau, J. (1997). Genetic and environmental influences on personality: A study of twins reared together using the self- and peer-report NEO-FFI scales. *Journal of Personality, 65*, 449-475.
- Rothbart, M.K. (1981). Measurement of temperament in infancy. *Child Development, 52*, 569-578.
- Rothbart, M.K. (1999). Temperament, fear, and shyness. In L.A. Schmidt & J. Schulkin (Eds.), *Extreme fear, shyness, and social phobia: Origins, biological mechanisms, and clinical outcomes* (pp. 88-93). New York: Oxford University Press.
- Rothbart, M.K., Ahadi, S.A., & Evans, D.E. (2000). Temperament and personality: Origins and outcomes. *Journal of Personality and Social Psychology, 78*, 122-135.
- Rothbart, M.K., & Bates, J. (1998). Temperament. In W. Damon (Series Ed.) & N Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 105-176). New York: Wiley.

- Rothbart, M.K., & Derryberry, D. (1981). Development of individual differences in temperament. In M.E. Lamb & A.L. Brown (Eds.), Advances in developmental psychology (Vol. 1, pp. 37-86). Hillsdale, NJ: Earlbaum.
- Sattler, J. (1988). *Assessment of children (3rd ed.)*. San Diego: J.M. Sattler Publishing.
- Schweder, R., & Bourne, E. (1984). Does the concept of the person vary cross-culturally? In R.A. Schweder & R.A. Levine (Eds.), *Culture theory: Essays on mind, self, and emotion* (pp. 158-199). New York: Cambridge University Press.
- Stafford, M.E. (1994). Validity of the STAR: Student Styles Questionnaire: Racial-Ethnic Comparisons (Doctoral dissertation, University of Texas) *Dissertation Abstracts*.
- Stafford, M.E., & Oakland, T.D. (1996). Validity of temperament constructs using the Student Styles Questionnaire: Comparison for three racial-ethnic groups. *Journal of Psychoeducational Assessment, 14*, 109-120.
- Steinmetz, J.E. (1994). Brain Substrates of Emotion and temperament. In J.E. Bates and T.D. Wachs (Eds.), *Temperament: Individual differences at the interface of biology and behavior*. (pp. 17-46). Washington, D.C.: American Psychological Association.
- Strelau, J. (1987). The concept of temperament in personality research. *European Journal of Personality, 1*, 107-117.
- Super, C.M., & Harkness, S. (1986). Temperament, development, and culture. In R. Plomin & J. Dunn (Eds.), *The study of temperament: Changes, continuities and challenges* (pp. 131-149). Hillsdale, NJ: Erlbaum.
- Tanzer, N.K. (2004). Developing tests for use in multiple languages and cultures: A plea for simultaneous development. In R. Hambleton, P. Merenda, & C.D. Spielberger (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Hillsdale, NJ: Lawrence Erlbaum.
- Teglasi, H. (1998). Introduction to mini-series: Implications of temperament for the practice of school psychology. *School Psychology Review, 27*, 475-478.
- Thayer, B. R. (1995). The relationship of temperament with respect to age, gender, and race/ethnicity in children and adolescents (Doctoral dissertation, University of Texas, 1995). *Dissertation Abstracts International, 56*, (10-A), 3893.
- Thomas, A., & Chess, S. (1977). *Temperament and development*. New York: Brunner/Mazel.
- Triandis, H.C., & Suh, E.M. (2002). Cultural influences on personality. *Annual Review of Psychology, 53*, 133-160.
- Tucker, L.R. (1951). *A method for synthesis of factor analysis studies* (Personnel Research Section Report No. 984). Washington, DC: Department of the Army.

- van de Vijver, F. & Leung, K. (1996). Methods of data analysis of comparative research. In J.W. Berry, Y.H. Poortinga, & J. Pandey (Eds.), *Handbook of Cross-Cultural Psychology, Volume 1: Theory and Method: Second Edition* (pp. 257-300). Pearson Allyn & Bacon.
- van de Vijver, F.J.R., Mylonas, K., Pavlopoulos, V., & Georgas, J.O. (2003). Methodology of Combining the WISC-III Data Sets. In J. Georgas, L.G. Weiss, F.J.R. van de Vijver, & D.H. Saklofske (Eds.), *Culture and Children's Intelligence: Cross-Cultural Analysis of the WISC-III*. Academic Press,
- van de Vijver, F.J.R., & Poortinga, Y.H. (1991). Testing across cultures. In R.K. Hambleton & J. Zaal (Eds.), *Advances in educational and psychological testing* (pp. 277-308). Boston, MA: Kluwer Academic Publishers.
- van de Vijver, F.J.R., & Poortinga, Y.H. (1992). Testing in culturally heterogeneous populations: When are cultural loadings desirable? *European Journal of Psychological Assessment*, 8, 17-24.
- van de Vijver, F.J.R., & Poortinga, Y.H. (2002). Structural equivalence in multilevel research. *Journal of Cross-Cultural Psychology*, 33, 141-156.
- Wachs, T.D. (1994). Fit, context, and the transition between temperament and personality. In C.F. Halverson, Jr., G.A. Kohnstamm, & R.P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 209-220). Hillsdale, NJ: Erlbaum.
- Watson, D. (2000). *Mood and Temperament*. New York: Guilford.
- Williams, M.Q., Williams, T.F., Qisheng, X., & Xuemei, L. (1992). A glimpse of the psychological types of mainland Chinese undergraduates. *Journal of Psychological Type*, 23, 3-9.

BIOGRAPHICAL SKETCH

Nicholas Frank Benson was born in Cooperstown, North Dakota. He lived there until he completed high school. He attended college in Minnesota and majored in psychology. In 1994, Nicholas received the degree of Bachelor of Arts from Moorhead State University (Moorhead, MN). He went on to pursue graduate studies in psychology at Illinois State University (Normal) where he received a Master of Science degree. In the fall of 1997, he enrolled at the University of Florida to study school psychology. He received a dual Master of Arts in Education/Educational Specialist degree in 2001. On completion of this degree, he began working as a school psychologist for Broward County Public Schools. When he married, Nicholas transferred to Miami-Dade County Public Schools and soon started a family. He plans to continue working with children. One day, he hopes to teach and research at the University level.