1-1-1995

Measurement and statistical models in the study of personality and intelligence.

Gregory J. Boyle
Bond University, Gregory_Boyle@bond.edu.au

Lazar Stankov
University of Sydney

Raymond B. Cattell
University of Hawaii

MEASUREMENT AND STATISTICAL MODELS IN THE STUDY OF PERSONALITY AND INTELLIGENCE

Gregory J. Boyle
School of Humanities & Social Sciences
Bond University
Gold Coast, Qld. 4229
Australia

Lazar Stankov
Department of Psychology
University of Sydney
Sydney, N.S.W. 2006
Australia

Raymond B. Cattell
Department of Psychology
University of Hawaii
Honolulu, Hawaii 96844

TABLE OF CONTENTS

Introduction

- Psychological Models: Historical Background
- Need for a Taxonomy of Psychological Constructs
- Role of Scientific Method in Elucidating Ability and Personality Structure
- Need for Multivariate Measurement and Experimentation
- Statistical vs. Clinical Interpretations of Individual Differences
- Cattellian Terminology and Philosophy of Research

Exploratory Factor Analysis: Appropriate Methodology

- Exploratory Factor Analytic Guidelines
- Sampling of Subjects and Variables
- Determination of the Appropriate Number of Factors
- Common Factor Model vs. Principal Components
- Oblique Simple Structure Rotation
- Testing the Significance of Derived Factors
- Role of Factor Analysis in Psychological Test Construction
- Aims and Scales of Factor Analytically Derived Instruments

Confirmatory Factor Analysis: Role in Validating Psychological Tests

- Exploratory-Confirmatory Factor Analytic Dualism
- Congeneric Factor Models
- Measurement versus Structural Models
- Goodness-of-Fit Indices
- Multitrait-Multimethod Matrices: Analyses of Covariance Structures

Structural Equation Modeling: Testing Measurement and Statistical Models

- Combination of Factor Analysis and Multiple Regression Analysis
- Advantages of Structural Equation Modeling
- Assumptions for Valid Use of LISREL
- Alternative Structural Modeling Packages
- Critique of Structural Equation Modeling Procedures

Psychometric Tradition in the Study of Personality and Intelligence
- Multivariate Psychometric Model
- Behavioral Specification Equations
- Factor Analysis of Intellectual Abilities
- Factor Analysis of Personality Traits
- Simplification of the Multivariate Psychometric Model
- Higher Stratum Cognitive Abilities
- Higher-Stratum Personality Dimensions
- Higher-Stratum Motivation and Mood-State Dimensions
- Varieties of Psychometric Measurement Media
- Motivation/Response Distortion in Personality Questionnaires
- Need for Objective Personality Test Construction
- Objective Motivation Measurement

Item Analysis Issues: Psychometric Properties in Personality and Intelligence Research
- Reliability: Stability vs. Dependability
- Item Homogeneity: Internal Consistency vs. Item Redundancy
- Item-Response Theory and Computerized Adaptive Testing
- Correlation Coefficients with Ordinal or Categorical Data
- Statistical Effect Size
- Generalizability Procedures
- Test Bias in Personality and Intelligence Research

Summary and Conclusions

MEASUREMENT AND STATISTICAL MODELS IN THE STUDY OF PERSONALITY AND INTELLIGENCE

by

Gregory J. Boyle, Lazar Stankov, and Raymond B. Cattell

Introduction

Psychological Models: Historical Background

Theorizing about personality and intelligence structure initially was limited to prescientific, literary and philosophical "insights" (cf. Howard, 1993). Among these early "psychological" approaches, Freudian psychoanalytic theory almost certainly has had the major influence on psychological thinking about human personality during the early 20th century--although psychoanalysis itself has now come under critical scrutiny (see Eysenck, 1985; Masson, 1990). Another prominent theorist was Murray who postulated several "Needs" such as: Abasement, Achievement, Aggression, Change, Cognitive Structure, Endurance, Nurturance, Order, Sentience, and Understanding. Likewise, Jung's Introversion-Extraversion theory has been influential. However, the comparatively subjective models of theorists such as Freud, Adler, Jung, Fromm, Erikson, Horney, Maslow, and Sullivan, must now be rejected as scientifically unacceptable. Around 1920, the emphasis changed from clinical premetric speculations to more quantitative and overtly experimental approaches, along with recognition of the ability and personality sphere concepts. The inadequacy of socioenvironmental explanations of personality, however, has been amply demonstrated by Zuckerman (1991) in his important *Psychobiology of Personality*. Personality is not solely the outcome of family and social conditioning. Eysenck (1991) has pointed out that these theories are essentially untestable. They are based on speculative or falsified deductions, and ignore virtually all the experimental and empirical research conducted this century.

Need for a Taxonomy of Psychological Constructs

Attempts to develop a taxonomy of cognitive abilities and personality traits have been based on the *factor analytic*[1] research of investigators such as Cattell, Comrey, Guilford, and Eysenck (cf. Brody, 1988, 1992; Carroll, 1991; Cattell, 1987a; Ceci, 1990). Investigators sought to measure empirically derived factors representing abilities and personality traits. The assumption of cross-situation stability of personality traits akin to that observed for cognitive abilities (changes in abilities occur throughout the lifespan), has been questioned by Mischel (e.g., 1984). However, this *situationist* philosophy has been thoroughly refuted and shown to be superficial (e.g., Boyle, 1985b, 1988c; Cattell, 1983; Conley, 1984; Eaves, Eysenck, & Martin, 1989; Eysenck, 1991; Eysenck & Eysenck, 1980; Kline, 1986). In summarizing two studies on this issue, Zuckerman (1991, p. 50) reported that "*Persons* accounted for almost the same percentage of variance in both studies (28-29%), and *persons* x *situations* interactions accounted for another significant portion of the variance (22-23%)." Clearly, empirical and experimental investigation of personality and intelligence necessitates the study of suitable intrapersonal psychological constructs--cognitive and personality traits, as Buss (1989) pointed out (see also chapter on facet theory approaches to domain definition by Most & Zeidner).

Role of Scientific Method in Elucidating Ability and Personality Structure

Application of scientific method to the study of personality and intelligence, has now emerged as the dominant mode of investigation (Kerlinger, 1986). A necessary, but not sufficient requirement of theoretically postulated "causal" relationships is correlation of the interrelated variables. Multivariate correlational analyses include procedures such as multiple regression analysis, path analysis, exploratory (EFA) and confirmatory (CFA) factor analysis, as well as the more sophisticated techniques of structural equation modeling (SEM)--(Bollen, 1989; Byrne, 1989; Byrne, Shavelson, & Muthén, 1989; Cuttance & Ecob, 1987). Psychometric measures provide an avenue for statistical hypothesis testing. Boyle (1988c) argued that measurement is the *sine qua non* of scientific investigation. Without quantitative measurements, it is simply not possible to test hypotheses, and consequently, to discriminate between competing theories or models of intelligence and personality.

That there is a complex interaction between intelligence and personality cannot be disputed (see Boyle, 1983b, 1987c, in press; Brody, 1992; Cantor & Kihlstrom, 1987; Cattell, 1987a; Goff & Ackerman, 1992). Conceptually, there may be analogies between personality and intelligence--both being construed as relatively enduring traits (Cattell, 1983). Intelligence test performance may be affected by personality attributes. Likewise, development of intellectual skills may be influenced by personality traits (Cattell, 1987a). Additionally, the interaction between intelligence and academic achievement is affected by personality factors. Anxiety can either interfere with, or facilitate performance, depending on the individual's competence or intelligence (Brody, 1992). Thus, in highly intelligent and/or competent individuals, heightened anxiety (e.g., under examination conditions) may enhance performance, whereas for less intelligent and/or less competent individuals, anxiety may have a debilitating impact on performance.

Need for Multivariate Measurement and Experimentation

In measuring personality and intelligence variables, there is a clearcut need for *multivariate* rather than univariate measurement (Boyle, 1991c; Horn, 1988; Nesselroade & Cattell, 1988). Intrapersonal psychological structure comprises a wide range of personality traits and cognitive abilities (Boyle, 1983b, 1987d; Cattell, 1979, 1980, 1982b, 1987a), so that multivariate measurement is necessitated. Experimental manipulation or therapeutic intervention may have significant effects on several psychological variables simultaneously, which univariate measurement is unable to monitor successfully (Boyle, 1985b). Multidimensional instruments for measuring intellectual abilities include the Comprehensive Analysis Battery or CAB (Hakstian & Cattell, 1982), Stanford-Binet Intelligence Scale or SB-IV (Thorndike, Hagen, & Sattler, 1986--cf. Boyle, 1989b), Wechsler Intelligence Scales (WAIS, WISC, WPPSI)--(Kaufman, 1990; Wechsler, 1991), British Ability Scales (Elliott, Murray, & Pearson, 1983), Kaufman Assessment Battery for Children or K-ABC (Kaufman & Kaufman, 1983), and the Woodcock-Johnson Psycho-Educational Battery-Revised (based on Gf/Gc theory--see Hessler, 1982; Woodcock & Mather, 1989). This multivariate experimental approach has been adopted extensively within the Cattellian school (cf. Boyle & Cattell, 1984; Stankov & Chen, 1988).

Statistical vs. Clinical Interpretations of Individual Differences

Clinical and psychiatric diagnoses are notoriously unreliable. Johnson (1986, p. 229) contended that diagnostic clinical ratings of interview data tend to be unreliable and low in validity. There is a clear need for diagnoses based on quantitative psychometric evidence, rather than on subjective observations (all too prevalent in various forms of psychotherapy). Countless papers on clinical vs. statistical (actuarial) prediction have supported the value of statistical-actuarial prediction. There is need for a multivariate quantitative (psychometric) approach to clinical practice. For instance, the Halstead-Reitan Battery has been one of the most useful tools for the clinical neuropsychological assessment of personality-intelligence interactions in relation to brain functioning. Whereas use of the Halstead-Reitan Battery has been based on a neuropsychological key approach over the past two decades (Russell, Neuringer, & Goldstein, 1970), the Luria-Nebraska Battery has been less popular (Boyle, 1986a).

A major problem in clinical neuropsychology has been inadequate incorporation of personality measures (including mood state and motivation dynamic trait measures) into research studies and applied clinical assessment. Neuropsychological test batteries have focused predominantly on cognitive aspects of brain functioning. Clearly, various forms of brain dysfunction are also associated with changes (from the normal) in non-ability intrapersonal characteristics such as personality, motivation and mood states (Powell, 1979). Zuckerman (1991, p. 169) stated that "*personality* depends on an intact, functioning brain...general psychiatric disturbance is proportional to the amount of brain destruction." These changes may have a profound effect on an individual's life, irrespective of cognitive functioning. There is therefore an urgent need to incorporate measures of non-ability intrapersonal variables into clinical neuropsychological assessment. Use of SEM approaches in the modeling of personality-cognitive interactions should greatly facilitate our understanding of underlying psychobiological mechanisms. However Zuckerman (p. 171) warned that too much emphasis is currently placed on animal models of human traits, and that the "paucity of human brain research, particularly on limbic systems, preclude(s) definitive statements now on the neuropsychology of personality traits....discovery that functional pathways in the brain are served by particular neurotransmitters has provided a new approach to identifying the circuitry involved in behavioral adaptations."

Cattellian Terminology and Philosophy of Research

Cattell saw the need for a taxonomy of psychological constructs (intellectual abilities, personality traits, dynamic motivation traits, and transitory mood states), somewhat akin to the periodic table of elements in chemistry. Therefore, he set out to discover (using the best available factor analytic techniques on comprehensive samples of variables and subjects) the major intrapersonal psychological dimensions. Using concise factor analytic solutions--even employing topological rotation over and above analytic methods, to achieve the highest level of simple structure possible (see section on exploratory factor analytic methods below), Cattell produced a taxonomy of abilities, traits, dynamics, and states. As his psychometric instruments have been constructed factor analytically, the scales therein are defined by discrete factors.

To avoid confusion over the meaning of these factors, Cattell coined several new terms to define his factors uniquely. Unfortunately, non-psychologists and even many research psychologists unfamiliar with Cattellian terminology have consequently been deterred because of an initial difficulty in knowing what he was talking about. Recently though, IPAT has simplified Cattell's terminology in the production of newer and more refined versions of his instruments, so that psychologists can no longer complain that the terminology is obscure and unnecessarily difficult to comprehend.

Exploratory Factor Analysis: Appropriate Methodology

Exploratory Factor Analytic Guidelines

Use of EFA in single-shot studies is potentially problematic (Guttman, 1992). EFA methods are driven by the idiosyncrasies of particular samples and therefore may serve to conflate theory. Romney and Bynner (1992) argued that EFA procedures produce "static" factors which are not sensitive to change. However, this criticism applies only to single-occasion R-factoring, whereas factoring of difference scores across measurement occasions (Dr-technique)--(see Boyle, 1987e), and factoring of an individual's scores over many repeated occasions (P-technique) has demonstrated the important role of dynamic motivation and transitory mood-state factors. Certain EFA procedures optimise the likelihood of obtaining a valid simple structure solution (see Boyle, 1988, pp. 742-745, 1993c; Cattell, 1978; Gorsuch, 1983; McDonald, 1985; and Mulaik, 1986). To obtain the best possible factor solution, a number of conditions should be satisfied.

Sampling of Subjects and Variables

It is necessary to strategically select variables to thoroughly cover the personality and ability domains. The general rule of thumb (cf. Gorsuch, 1988) is that a minimum 10 subjects per variable is required to obtain accurate factor pattern solutions. Even with 300 subjects, the appropriate factor solution is obtained in only 50% of cases. According to Cuttance (1987, p. 243), "MacCallum (1985) investigated the process of the exploratory fitting of models in simulated data...*only about half of the exploratory searches located the true model...in samples of 300 observations...and his success rate in smaller samples (N = 100) was zero* [italics added]...the probability of locating the correct model by exploratory methods when sample data are used is even less..."

Consequently, we have to assume that many of the EFA studies reported in the psychological literature have been flawed due to inadequate sampling of variables and subjects, particularly in studies of multidimensional personality inventories where many variables are involved (cf. Cudeck & Henly, 1991). For example, in a recent study of personality-intelligence relationships, Goff and Ackerman (1992) undertook several EFA analyses based on the intercorrelations of combined personality and ability measures, using a sample of only 138 subjects. In view of MacCallum's findings, one would expect the Goff and Ackerman factor solutions to be unreliable and of dubious validity. Indeed, Goff and Ackerman's solutions did not satisfy simple structure requirements as shown by inadequate ±.10 hyperplane counts (see below). Aside from utilizing appreciably larger samples (500 subjects or more), another avenue is to take a "two-handed" approach, wherein the factor models derived from exploratory methods are subjected to goodness-of-fit testing using CFA methods (e.g., Boyle, Borg, Falzon, & Baglioni, 1993).

Determination of the Appropriate Number of Factors

Every subsequent step in an EFA analysis will be adversely affected if a less-than-optimal number of factors is extracted. The decision as to number of factors is influenced by several considerations, including various psychometric and objective tests, and the degree to which simple structure is attained. Empirical research (Hakstian, Rogers, & Cattell, 1982) has demonstrated the utility of the Scree test. As compared with the Kaiser-Guttman (K-G) eigenvalues greater than unity criterion, the Scree test is more accurate when there are fewer than about 20, or more than 40 to 50 variables (Child, 1990). The scree test has been automated both by Barrett and Kline (1982), and separately by Gorsuch and Nelson (1981)--(see Gorsuch, 1983). Use of these algorithms removes the subjectivity in determining the relevant "scree break." Objective tests for determining the number of factors include, for example, the Very Simple Structure (VSS) method, the asymptotic chi-square statistic, Bartlett's test of equality of the last p - m eigenvalues, and Velicer's minimum average partial (MAP) test (cf. Loehlin, 1990; Velicer & Jackson, 1990a,b). The rotated factor pattern provides a final index of the accuracy of number of factors. The $\pm.10$ hyperplane count (percentage of variables with trivial factor loadings) provides a quantitative index of the extent of simple structure (Boyle, 1993d; Boyle & Stanley, 1986; Cattell, 1978; Gorsuch, 1983). Use of various tests in conjunction with criteria for over- and underextraction, and consideration of hyperplane counts facilitate determination of the appropriate number of factors.

Common Factor Analysis versus Principal Components

Principal components analysis (with unities in the leading diagonal of the correlation matrix) artificially inflates factor loadings, due to spurious common factor variance (Comrey & Lee, 1992). Principal components analysis is mathematically elegant, but the psychological interpretability of the derived components may be less than optimal. Iteration of communality estimates accords with the common factor model. When the number of variables is greater than about 20, iteration actually makes little difference to the factor solution. Gorsuch (1990, pp. 35-36) suggested that at least two to three iterations need be carried out.

Convergence of communalities proceeds rapidly for well defined problems, where a factor solution is reliable. As Velicer and Jackson (1990a) have pointed out, poorly defined factors loading on only a few variables (with small loadings), and/or extraction of an inappropriate number of factors, inevitably results in an excessive number of iterations required to reach convergence. Use of principal components analysis provides no indication of the reliability of the solution, whereas the number of iterations in common factor analysis provides a direct (inverse) index of factor reliability. Principal components analysis is a poor substitute for common factor analysis (Cattell, 1978; Gorsuch, 1983; McArdle, 1990; McDonald, 1985).

However, there are still some proponents of the short-cut principal components (PCA) method (e.g., Schönemann, 1990; Velicer & Jackson, 1990a). They have argued that PCA avoids the problem of factor indeterminacy, and is computationally more efficient. Their argument, based on expediency and computational speed, is hardly relevant given modern computing facilities. Moreover, Mulaik (1990, p. 54) asserted that, the indeterminacy associated with common factor model is really just an example of the pervasive indeterminacy that exists throughout all science (cf. Rozeboom, 1990). Snook and Gorsuch (1989) reported that simulation studies show that PCA gives discrepant results, when the number of variables in the analysis is low (cf. Widaman, 1990). They also reported that component loadings are systematically inflated, as compared with factor loadings. Likewise, Bentler and Kano (1990) pointed out that common factor analysis is preferable to the PCA approach. Gorsuch (1990, p. 39) concluded that use of common factor analysis "recognizes we have error in our variables, gives unbiased instead of inflated loadings...use of components is primarily the result of decisions made when there were problems computing common factor analysis which no longer exist and the continuation of its being a ready default on computer programs designed during an earlier era."

Oblique Simple Structure Rotation

In accord with Thurstone's simple structure principles (see Child, 1990, pp. 48-49), a unique oblique factor pattern solution is usually desirable. Only when simple structure is achieved, is it possible for the resultant factors to have the status of *causal* determinants (Kline, 1980), although causality cannot be inferred solely on the basis of correlational evidence. Use of orthogonal rotation often fails to achieve simple structure. In fact, an oblique rotation to maximum simple structure will stop at the special orthogonal position in the event that uncorrelated factors are actually warranted. Maximum simple structure is often not attained with analytic oblique rotation alone. In general, the higher the hyperplane count, the better is the simple structure of the factor solution, with ±.10 hyperplane counts of at least 65-70% suggesting an adequate attainment of simple structure. Thus, in Goff and Ackerman's (1992) study, an orthogonal factor solution exhibited a hyperplane count of only 20.0%, revealing its invalidity. A corresponding oblique factor pattern solution gave a hyperplane count of 58.5%, in accord with the general superiority of oblique versus orthogonal solutions.

It may be necessary to undertake additional topological rotation via Rotoplot (Cattell, 1978). Studies have shown the efficacy of Rotoplot (Cattell, 1978) to improve the resultant ±.10 hyperplane count (Boyle & Stanley, 1986). Nevertheless, the increase is often so slight as not to warrant the extra expenditure of time and effort. Measurement noise due to idiosyncrasies of particular samples suggests that the search for simple structure in single sample data may be problematic, and less important than replication and crossvalidation of results. Although statistical software exists for the easy use of Rotoplot (e.g., Brennan & Nitz, 1986), more important, is the need to test the goodness-of-fit of proposed factor models via CFA methods.

Testing the Significance of Derived Factors

One can test the significance of derived factors, using the Kameoka and Sine tables (in Cattell, 1978). Boyle (1988c) demonstrated that these tables are overly conservative in failing to attribute significance to recognizable factors when other criteria clearly show such factors to be meaningful. Less restrictive use of these tables could provide useful information on the significance of factors derived from exploratory methods. Ideally, the invariance of factors (cf. Byrne, 1988) should be checked across different samples at both primary and second-stratum factor levels. One approach is to employ Cattell's (1978) congruence and salient variable indices, which provide a more accurate indication of factor invariance than does a simple correlational analysis of factor loadings. Perusal of published factor analytic research in psychology and the social sciences reveals that this level of crossvalidation recommended by Cattell has rarely been attempted, let alone achieved.

Role of Factor Analysis in Psychological Test Construction

Use of factor analysis provides important evidence as to construct validity, but this evidence alone is insufficient. In addition to factor validity, predictive validity evidence is essential (e.g., O'Toole & Stankov, 1992). Factor validity is a necessary precondition, which at best, is suggestive of construct validity (also see section on the examination of MTMM data via CFA techniques). In general, EFA methods support a hierarchical model for both personality traits and intellectual abilities. However, Romney and Bynner (1992) argued that EFA cannot reveal a simplex structure, wherein there is a linear ordering of tests--amounting to a conceptual limitation of the common factor model. They suggested that cognitive abilities might be explained more adequately in terms of a dynamic split-simplex model, comprising a linear ordering of abilities, rather than resorting to explanations in terms of an underlying common factor (see section below, on the factor analysis of abilities). However, Stankov and Crawford (1993) argued that complexity of a series of cognitive tasks is revealed by the size of their loadings in relation to the general factor, which in turn, is defined by these tasks and other cognitive measures. This pattern of loadings may not be related to the linear ordering of tasks, *per se*.

Briggs and Cheek (1986, p. 106) recommended routine application of factor analysis in the construction and validation of new personality scales (factor analysis is superior to the superficial approach of cluster analysis--Boyle, 1985b; Cattell, 1978; McArdle, 1984). Factor analysis is an important aspect of construct validation. For example, Boyle (1987a) administered the Eight State Questionnaire (8SQ) and the Differential Emotions Scale (DES-IV) to a sample of 212 undergraduate students on two occasions, and factor analyzed the difference scores (dR-factoring)--(cf. Boyle (1987e). Using an iterative principal factoring procedure and oblique simple structure rotation, four higher-order mood-state dimensions emerged. Results suggested that two broad mood-state dimensions are measured within each instrument. The first DES-IV factor loaded on Guilt, Sadness, Hostility, Fear, Shame, and Shyness, representing negative emotionality akin to Eysenck's Neuroticism dimension. The second DES-IV dimension was a bipolar factor which contrasted Interest, Joy, and Surprise with Anger, Disgust, Contempt, and Guilt. The first 8SQ factor contrasted positive (Extraversion, and Arousal) emotions with negative (Depression, and Fatigue) states, whereas the second factor loaded on several Neuroticism states (Anxiety, Stress, Depression, Regression, and Guilt). Thus, each instrument could be simplified internally, enabling more efficient measurement of central mood states. In another example, Boyle (1988a) administered the Profile of Mood States (POMS) and the 8SQ to 289 undergraduates. Higher-order scale factoring of the combined instruments revealed four major state dimensions (Neuroticism, Hostility/Anger, Vigor, and Extraversion vs. Fatigue-Arousal). These findings provided evidence on the internal structure of the two instruments, showing the relationship of higher-stratum dimensions to primary factors. However, in ascertaining the construct validity of an instrument, factor analysis represents only one approach, along with correlational and experimental analyses.

Aims and Scales of Factor Analytically Derived Measurement Instruments

Although requiring further refinements, the CAQ extends measurement into the abnormal personality trait domain (cf. Boyle, 1990b; Guthrie, 1985). Part 1 measures the usual 16PF factors, plus another six higher-stratum dimensions (see section on higher-order factors below), whereas Part 2 measures 12 separate factor analytically derived psychopathology scales (Kameoka, 1986), and *at least* five major abnormal dimensions at the second-stratum level (Boyle, 1987d). The clinical factors are labelled D1 (Hypochondriasis), D2 (Suicidal Depression), D3 (Agitation), D4 (Anxious Depression), D5 (Low Energy Depression), D6 (Guilt and Resentment), D7 (Boredom and Withdrawl), Pa (Paranoia), Pp (Psychopathic Deviation), Sc (Schizophrenia), As (Psychasthenia), and Ps (Psychological Inadequacy).

Two limitations of the current version of the CAQ are (1) insufficient numbers of items in Part 1 (which has only eight items in each of the 16 scales, although supplementation with other forms of the 16PF is a viable option), and only 12 items per scale in Part 2; and (2) the factor structure of the abnormal trait sphere (CAQ Part 2) needs to be refined and crossvalidated using both exploratory and confirmatory factor analytic procedures on independent samples. The factor analytic basis of the CAQ is deficient since some 45 separate factor analytic studies of subsets of the combined MMPI and depression item pool were undertaken, rather than a single factoring of item parcels. Emergence of seven separate depression factors is an artifact due to inclusion of an excessive number (200-300) of depression items in the factor analyses, over and above the MMPI item pool. Consequently, Part 2 of the CAQ has dubious factor validity. Kline (1993b) has discussed some of the limitations of non-factored scales such as the criterion-keyed MMPI. Scales which are not factor valid, cannot clarify the causal mechanisms involved in psychopathological processes. In assessing the causal determinants of personality and intelligence, factor-valid scales are undoubtedly a great asset.

In a study of the interbattery correlations of the 14 scales in the High School Personality Questionnaire or HSPQ--a downward extension of the 16PF and the 20 CAB ability measures, no fewer than 50 out of 280 correlations were significant. Only 14 of these correlations would have been expected to be significant by chance alone (at the $p<.05$ level). While the ability and personality domains are conceptually distinct, it is clear that artistic, mathematical and verbal skills are associated with various personality traits (see Cattell, 1987a). What are often thought of as different qualities of ability are probably complex combinations of cognitive abilities and personality traits. Studies with the 16PF and HSPQ have shown a significant increase in prediction over that based on intelligence tests alone (e.g., Boyle, 1983b; Boyle, Start, & Hall, 1989). Cattell (p. 480) reported an average 42% increase by including personality in addition to cognitive ability measures alone. There can be little doubt about the combined role of personality and intelligence in influencing academic learning outcomes.

Confirmatory Factor Analysis: Role in Validating Psychological Tests

Exploratory-Confirmatory Factor Analytic Dualism

A *two-handed* approach to factor analysis of the personality and ability domains is desirable. Exploratory (EFA) and confirmatory (CFA) factor analyses should be carried out on independent samples, and both sets of analyses crossvalidated (cf. Bryne, Shavelson, & Muthén, 1989). Results from an exploratory analysis enable an empirical test (via CFA) of empirically derived models. Confirmatory methods are conceptually driven, wherein models are tested statistically against empirical data for their "goodness-of-fit." Confirmatory methods enable statistical model testing, unlike the traditional data-driven, exploratory approaches (cf. Anderson, 1987; Bentler, 1989; Breckler, 1990; Muthén, 1988). In EFA, the latent variable structure usually is unknown, and the focus is on discovering the main factors underlying observed variables. In contrast, CFA is applicable when the latent variable structure has already been suggested on theoretical, empirical, or other grounds (Byrne, 1989; Marsh & Bailey, 1991). Nevertheless, CFA can produce discrepant results (Bagozzi & Yi, 1990; Millsap, 1990; Williams, Cote, & Buckley, 1989).

A common misconception is that EFA is now superseded by CFA. This view could not be further from the truth. The two procedures are complementary, not competing methodologies (Bentler, 1988). EFA is undertaken to map out the factor structure within a domain, while CFA is applied to an independent sample to test the fit of the factors previously located (cf. Bentler, 1990a; Cuttance & Ecob, 1987; MacCallum, 1986; McDonald & Marsh, 1990). This dual approach to elucidation of factors on the one hand, and their verification on the other, is the desirable way to proceed.

Congeneric Factor Models

One of the best approaches is to undertake CFA, including congeneric one-factor analyses, via PRELIS (Jöreskog & Sörbom, 1988), followed by LISREL (Jöreskog & Sörbom, 1989). Use of PRELIS is important particularly if any of the variables are noticeably skewed or kurtotic, and when dealing with categorical or ordinal data (as indicated above). A major use of CFA is in the validation of psychological tests. CFA procedures enable assessment of the factor structure of an instrument, and also the appropriateness of the item content of each scale. Boyle (1990a, 1991c, 1992a; Boyle & Fabris, 1992) has undertaken confirmatory analyses of the SB-IV, 8SQ, Menstrual Distress Questionnaire or MDQ (Moos, 1985), and Holland's (1985) Self-Directed Search (SDS), respectively. Likewise, Byrne (1989) has carried out extensive confirmatory factor analyses of the Self-Description Questionnaire (see Boyle, 1993c). Many of the extant instruments have a multidimensional scale structure (e.g., 16PF, CPI, MMPI, MDQ, 8SQ, POMS, DES-IV). *What is now needed is a systematic application of confirmatory methods, to verify the claims of test authors regarding the dimensionality of existing* *personality and intelligence test instruments.*

Measurement versus Structural Models

CFA involves the *measurement* part of the full structural equation model (which comprises both *measurement* and *structural* submodels--cf. Cuttance & Ecob, 1987). CFA is applied to either an all-X (exogenous) or all-Y (endogenous) model. According to Byrne (1989, p. 8), specifications are made with respect to "(a) The number of factors (ξ's or η's). (b) The number of observed variables (x's or y's). (c) Relations between the observed variables and the latent factors (λ_xs or λ_ys). (d) Factor variances and covariances (Φ). (e) Error variances (and possibly covariances) associated with the observed variables (Θ_δ or Θ_ε)." The measurement model (Jöreskog & Sörbom, 1989) is expressed algebraically as:

$$x = \Lambda_x\xi + \delta \quad \text{and} \quad y = \Lambda_y\eta + \varepsilon \qquad \text{--------------} \ (1)$$

wherein the observed variables are represented by the x's or y's, and the latent variables by the ξ's or η's, respectively. The δ and ε values represent the vector of measurement errors. The corresponding equation for the covariance matrices among the x variables is:

$$\Sigma = \Lambda\Phi\Lambda' + \Theta_\delta \qquad \text{---------------} \ (2)$$

wherein Λ represents the matrix of latent trait loading, Φ stands for the matrix of covariances between the latent traits, and Θ_δ represents the matrix of error variances and covariances. A similar equation pertains for the covariation matrices among the y variables. The full LISREL structural equation system among the η and ξ latent variables is represented by:

$$\eta = B\eta + \Gamma\xi + \zeta \qquad \text{-----------------} \ (3)$$

The vectors η and ξ represent the latent dependent and independent variables, whereas B(m x m) and Γ(m x n) represent the coefficient matrices, and ζ represents a random residual vector (involving random disturbance estimates, and errors in equations)--(see Jöreskog & Sörbom, 1989, p. 3). Thus the full LISREL model

(Bollen, 1989) incorporates three separate equations (covering the measurement models for *x* and *y*, and the structural equation model).

Goodness-of-Fit Indices

The goodness-of-fit (GFI) index assesses the fit of proposed models to empirical data sets. The GFI, which ranges from zero through 1.0, provides an estimate of the variance/covariance accounted for by models. The Adjusted Goodness-of-Fit (AGFI) index (which adjusts the GFI for the number of degree of freedom) and the Root Mean Square Residual (RMR) are two of the most important indices to consider. The RMR provides an estimate of the discrepancy between the predicted and observed covariance matrices. Better models have AGFI indices close to one, and RMR indices close to zero (values less than 0.05)--(see Bryne et al., 1989). According to Cuttance and Ecob (1987, p. 260), "models with an AGFI of less than .8 are inadequate...*acceptable models would appear to have an AGFI index of greater than .9* [italics added]."

Thus, in Boyle's (1990a) study of the SB-IV dimensionality, the scale intercorrelations for all 5,013 subjects (reported in the Technical Manual) were subjected to a CFA analysis via PRELIS/LISREL. The initial two-stage least squares solution served as the starting point for the maximum likelihood (ML) estimation. The resulting AGFI was .87 (RMR = .05). The Total Coefficient of Determination was .99 for the four SB-IV Area dimensions (see Thorndike et al., 1986). Congeneric (one-factor CFA) analyses supported the four Area dimensions. Thus, for Verbal Reasoning, the AGFI was .89 (RMR = .03); for Abstract/Visual Reasoning, the AGFI was .99 (RMR = .01); for Quantitative Reasoning, the AGFI was .99 (RMR = zero); while for Short-Term Memory, the AGFI was .96 (RMR = .02).

Boyle's (1991d) CFA analysis of the 8SQ was undertaken on the polychoric item intercorrelations (cf. Poon & Lee, 1987), computed via PRELIS across all 1,111 subjects. The resulting AGFI was .71 (RMR = .10), indicating an inadequate fit of the eight-factor model (Anxiety, Stress, Depression, Regression, Fatigue, Guilt, Extraversion, and Arousal). Congeneric analyses provided stronger support for the purported subscale structure (mean AGFI was .93; mean RMR = .04). Exogenous latent trait covariances revealed some measurement overlap of scales. The CFA item analysis of the MDQ on a sample of 369 female undergraduates (Boyle, 1992a) resulted in an AGFI of .87 (RMR = .06), suggesting a reasonable fit of the proposed eight-factor model. Congeneric analyses suggested that some MDQ scales are stronger than others (mean AGFI = .85; mean RMR = .05).

Boyle and Fabris (1992) undertook a CFA on the Self-Directed Search or SDS (five variables for each RIASEC theme--Realistic, Investigative, Artistic, Social, Enterprising, and Conventional), on a sample of 401 subjects. The AGFI of .75 (RMR = .08), failed to support the postulated RIASEC model. Congeneric results revealed an inadequate fit for the Realistic theme (AGFI = .78; RMR = .09). For the other RIASEC themes, the mean AGFI was .89 (mean RMR = .05). Covariances between exogenous latent traits suggested considerable measurement overlap between RIASEC categories.

Boyle et al. (1993) administered a sources of stress inventory to elementary school teachers in Malta. The group of 710 full-time teachers was randomly split. An EFA on the first subsample, produced a five-factor oblique solution. Factors were labeled: (1) Workload; (2) Professional Recognition; (3) Student Misbehavior; (4) Classroom Resources; (5) Relations with Colleagues (see Borg, Riding, & Falzon, 1991, for item details). The factor solution exhibited a $\pm$.10 hyperplane count of 54%, indicating better simple structure than for a four-factor solution (hyperplane count 45%). CFA on the second subsample, supported the five-factor model (AGFI = .81; RMR = .076). A fully-trimmed non-recursive model yielded an AGFI of .80 (RMR = .11), suggesting an acceptable fit. Congeneric factor analyses also provided strong support for each of the hypothesized factors (mean AGFI = .96; mean RMR = .02). Incremental fit indices rho and PNF12 (Mulaik, James, van Alstine, Bennett, Lind, & Stilwell, 1980) enabled comparison of the various models.

What is now needed is the testing of new models of personality and intelligence using SEM techniques, wherein the latent traits are regressed onto each other. Such an approach should throw light onto the nature of ability-personality interconnections, and interactions. SEM offers much hope for the development of a far more sophisticated understanding of such psychometric interrelationships than currently exists. With exploitation of SEM methods to their full extent (noting limitations alluded to by Breckler, 1990), psychometrics will undoubtedly become one of the most important and exciting fields of psychological research.

Multitrait-Multimethod Matrices: Analyses of Covariance Structures

An innovative application of CFA has been in the modeling of multitrait-multimethod data Cole & Maxwell, 1985). Byrne and Goffin (in press) have discussed new approaches to the investigation of multitrait-multimethod matrices (MTMM) involving analyses of covariance structures. These models include Jöreskog and Sörbom's (1988) general confirmatory factor analytic model (CFAGEN), Marsh's (1989) correlated uniqueness CFA model (CFACU), and Browne's (1984) composite direct product (CDP) model. According to Byrne and Goffin, the general CFA model enables "(a) an explanation of the MTMM matrix in terms of underlying latent constructs, rather than observed variables, (b) the evaluation of convergent and discriminant validity at the matrix, as well as at the parameter level, (c) the testing of hypotheses related to convergent and discriminant validity, and (d) separate estimates of variance due to traits, methods, and error, in addition to estimated correlations for both trait and method factors." Schmitt and Stults (1986) have critically reviewed the strengths and weaknesses of traditional MTMM approaches to construct validation. Byrne and Goffin have listed several major difficulties with the traditional MTMM approach. They suggested that researchers should estimate all three of the above covariance structure models, accepting the best fitting one. As they also pointed out (p. 27), "The imminent availability of fit indices for which confidence intervals have been statistically derived (Steiger, 1989; Browne, 1990) holds great promise for the assessment of such competing models." A comprehensive review of the problems associated with application of CFA and MTMM approaches was provided by Marsh (1989; Marsh & Bailey, 1991). Recently, a method for undertaking multiple group CFA analyses (using the UniMult program) has been devised by Gorsuch (1991), which should be useful in the modeling of MTMM matrices.

Structural Equation Modeling: Testing Measurement and Statistical Models

Combination of Factor Analysis and Multiple Regression Analysis

Structural equation modeling or SEM (Anderson & Gerbing, 1988; Cuttance & Ecob, 1987; Martin, 1987) involves the simultaneous application of factor analysis wherein the latent traits (factors) load on the observed variables (measurement model), and multiple regression analysis of the latent traits on each other (structural model)--(Byrne, 1988). McArdle (1984) pointed out that contemporary modelers can learn much from Cattell's structural modeling endeavors. SEM combines the factor (measurement) and path (structural) models into a single model, wherein each latent trait (factor) is regressed onto the others. It is assumed that for each latent trait, the residual and error terms do not correlate either with the factor or each other. In some instances one might question the validity of this assumption. SEM should facilitate scientific hypothesis testing, in contrast to exploratory approaches, which historically, have only served to conflate theory, rather than discriminating between competing hypotheses.

Boyle (1993b) investigated interrelationships among 8SQ mood states and menstrual cycle symptoms (measured via the MDQ on a sample of 370 undergraduate women. Factor analytic (EFA) results suggested that 8SQ states loaded on two separate factors--one involving neurotic states Anxiety, Stress, Regression, and Guilt; the other contrasting Depression and Fatigue with Extraversion and Arousal. Likewise, the MDQ scales separated into two distinct factors--one loading on psychological scales Negative Affect, Impaired Concentration, and Behavior Change; the other loading on physical symptoms Autonomic Reactions, Pain, and Water Retention. A LISREL SEM analysis tested both recursive and non-recursive models. For the non-recursive model, all parameters were identified, and the AGFI was .98 (RMR = .04), suggesting a reasonable fit to the data. This model suggested that psychological and physiological states and symptoms interact in a complex manner.

Advantages of Structural Equation Modeling

SEM has the advantage of being able to estimate the magnitude of error terms, unlike the older approach of path analysis which relied solely on multiple regression procedures, and which simply assumed that error terms are zero (cf. Kaplan, 1990). Structural modeling allows statistical testing of the fit of hypothesized models against actual empirical data sets (Bentler, 1990; Connell, 1987; Tanaka, 1987). Variance associated with measurement noise can be partialled out by removing variables with excessive error and which contribute little valid variance ("noisy" variables). Perusal of standardized regression equations associated with the LISREL two-stage least squares estimation procedure, suggests which variables should be deleted. This attenuation of "measurement noise" facilitates testing of postulated models. Structural modeling packages (such as LISREL/COSAN/EQS) should be used to investigate the *causal* influence of personality and intelligence variables to behavioral outcomes. Another recent advance has been in multilevel modeling packages (such as ML3--see Prosser, Rasbash, & Goldstein, 1991) which, when integrated into SEM packages such as LISREL, should facilitate a much more sophisticated analysis of psychometric models of personality and intelligence.

Assumptions for Valid Use of LISREL

Several conditions must hold for valid use of LISREL in testing the fit of proposed models (cf. Bollen, 1989; Cuttance & Ecob, 1987; Hayduk, 1987; Marsh, Balla, & McDonald, 1988; Romney & Bynner, p. 14). Parameters of the model should be determined uniquely--only one solution to the set of simultaneous equations should be found ("identified" model). Second, model parameters should be estimated via an iterative procedure such as maximum likelihood (ML), or other methods such as weighted least squares (WLS), or generalized least squares (GLS). Third, given the assumption of multivariate normality, the residual matrix approximation to zero is tested by a likelihood ratio (chi-square) test. Unfortunately, this test is sample-size sensitive, so that with large samples, virtually all proposed models are rejected, but to minimize sampling bias, large samples are desirable (Cudeck & Henley, 1991). Fourth, modification indices for parameters constrained to zero indicate the reduction in chi-square values when parameter constraints are released. Fifth, for non-continuous variables, computation of Pearson product-moment correlation coefficients may result in significant bias. PRELIS enables computation of polychoric and polyserial correlation coefficients, as required.

Alternative Structural Modeling Packages

Other structural modeling packages include LISCOMP used with categorical data (Muthén, 1988), COSAN used with interval data (McDonald, 1985), EQS (Bentler, 1985, 1989), and ProcCALIS (Hartmann, 1990). Statistical testing of proposed models enables some assessment of the causal determinants of various intellectual and personality variables on behavioral outcomes (cf. Biddle & Marlin, 1987; Mulaik, 1987). SEM merges CFA, multiple regression analysis, and path analysis into a single model, and provides a means of discriminating between competing hypotheses and models, in accord with scientific method. Nevertheless, there are limitations of SEM. However, as Breckler (1990) pointed out, in many of the published applications there are serious flaws. Even though fit of the desired model is identical for a large number of possible equivalent models, this is seldom acknowledged.

Critique of Structural Equation Modeling Procedures

Several potential difficulties in the application of SEM techniques have been discussed comprehensively by Breckler (1990). Likely problems include (1) computation of feasible parameter estimates when certain parameters are not identified fully; (2) use of the sample-size dependent chi-square test; (3) interpretation of the root mean square residual (RMR) index in covariance units; (4) unrecognized equivalent models which are not tested for their fit; (5) tendency toward reification of latent variables; (6) inaccurate modification indices; (7) drawing causal inferences when the data only provides suggestive relationships between latent variables. According to Bentler (1988, p. 3),

> ...the generative theory may be inappropriate, key variables may be omitted, samples may be biased, ambiguity may exist about causal ordering, measurement may be unrepresentative and inadequate, sampling of variables may be arbitrary, time lags for effects may be unknown, and the meaning of latent variables may be obscure. Furthermore, models may not be tested against independent data...inappropriate emphasis appears on confirmatory rather than exploratory data analysis; and...SEM tends to be applied subjectively and in a post-hoc manner. Key structural assumptions...linearity and additivity of relations, and the statistical assumptions of independent, identical distributions of observations, random sampling...large samples and multivariate normality, may not be

plausible..."

Advantages of SEM techniques have been overemphasized, and the validity of proposed structural models is directly related to the adequacy of the data and the sample employed. Testing of competing models will always be plagued by inadequate empirical data sets (e.g., data collected from rather unreliable measurement instruments). In fact, application of CFA methods to the validation of psychological instruments may be problematic. Often when there are more than three items per scale, the CFA analysis produces suboptimal GFI and AGFI indices. This is related to the unreliability of individual items ("noisy items") comprising such personality instruments or intelligence tests. Even though SEM techniques provide new opportunities for advances in psychological knowledge, these techniques are not a panacea for extracting meaning out of "sloppy" data. The age-old problem of "garbage in--garbage out" (GIGO) applies equally to all statistical methods, including SEM approaches.

Psychometric Tradition in the Study of Personality and Intelligence

Multivariate Psychometric Model

The multivariate psychometric model is an extension of the traditional trait model into other intrapersonal psychological domains. It is based on the mathematico-statistical technique of factor analysis, which determines the major variables for inclusion within the model. The most elaborate development of the psychometric model for behavioral prediction has been by Cattell and his colleagues, with each of the factor analytically elucidated ability, trait, dynamic, and state dimensions contributing to various versions of the "behavioral specification equation" (e.g., Cattell, 1979/1980; 1983). Kline (1980) pointed out that: (1) factors may in some instances have causal properties; (2) they represent the most important variables, provided variables and subjects are comprehensively sampled; (3) rotation to oblique simple structure facilitates determinate solutions; (4) maximization of the ($\pm$.10) hyperplane count (Cattell, 1978) results in simple structure solutions; (5) marker variables should be included. Kline contended that the psychometric model comprises the most important simple structure factors which have emerged in each of the ability, personality, motivation dynamic, and mood-state domains. Unfortunately, many factor analytic studies (e.g., Costa &

McCrae, 1991; Goff & Ackerman, 1992; McCrae & Costa, 1987, 1989; Zuckerman, Kuhlman, & Camac, 1988) have been plagued by failure to attain maximum simple structure, as advocated by Thurstone (cf. Child, 1990).

Behavioral Specification Equations

Kline (1980) concluded that the Cattellian psychometric model enables valid predictions of behavior; shows the inadequacy of the situationist argument; and facilitates systematic studies in basic and applied psychological research. Cattell's behavioral specification equations (see simplified Equations 4 and 5), differ in their complexity, and combine the action of cognitive abilities (A), normal and abnormal personality traits (T), dynamic motivation traits (D), and transitory mood states (S). By definition, for individual *i*, the '*a*'s represent behavioral outcomes of the response/performance *j*, whereas the *b*'s represent factor loadings/behavioral indices, as a function of the focal stimulus *h*, and the ambient situation *k*. This quantitative predictive approach is useful in showing the important role of various intrapersonal psychological variables. Clearly, there is a complex interaction between psychological and situational variables in influencing behavioral outcomes.

$$a_{hijk} = \Sigma\, b_{hjkw}\,\mathbf{A}_{wi} + \Sigma\, b_{hjkx}\,\mathbf{T}_{xi} + \Sigma\, b_{hjky}\,\mathbf{D}_{yi} + \Sigma\, b_{hjkm}\,\mathbf{S}_{mi} \quad \text{----------------- (4)}$$

With respect to the first-order personality by intelligence interaction, the multiplicative term is shown in Equation 2, below, in simplified form.

$$\Sigma\Sigma\, b_{hjkwx}\,\mathbf{A}_{wi}\,\mathbf{T}_{xi} \quad \text{------------------- (5)}$$

A detailed presentation and discussion of more sophisticated versions of these prediction equations, including both multiplicative and non-linear terms, is provided in Boyle (1988c). However, while specification of such behavioral prediction equations is theoretically justified, in practice, it is "well nigh impossible" to empirically quantify the various factor loadings in most instances.

Factor Analysis of Intellectual Abilities

Since intelligence has been viewed as directly related to efficient neurological functioning, measures such as reaction time (RT) and visual acuity have been regarded as appropriate. This line of research was extended by Spearman, who examined rank-difference intercorrelations (see Boyle & Langley, 1989) between ability measures (Brody, 1992; Jensen, 1991; Snow, Killonen, & Marshalek, 1984). Thurstone's development of multiple factor analysis enabled the structural dimensionality of abilities to be elucidated within the constraints of a hierarchical model (cf. Carroll, 1984; Cattell, 1982b; Guilford, 1985; Horn & Stankov, 1980; Messick, 1992; Stankov & Horn, 1982).

Thurstone delineated several primary ability factors which he labelled, Spatial, Perceptual, Numerical, Verbal Relations, Word Fluency, Memory, and Induction (subsequently extended to 20 primary abilities, as measured in the CAB (see Hakstian & Woolsey, 1985; Kline & Cooper, 1984). At first sight, it appeared that Spearman's general ability factor (**g**), and Thurstone's primary mental ability factors were incompatible (Carroll, 1991; Kranzler & Jensen, 1991). Cattell (1982b, 1987a) resolved this apparent discrepancy by factor analyzing Thurstone's primary mental ability intercorrelations, and found that at the higher-order level, general factors (Gf and Gc) emerged (cf. Boyle, 1988b; Stankov, 1978, 1983, 1986; Stankov & Chen, 1988; Stankov, Horn, & Roy, 1980). Hence, both Spearman's and Thurstone's findings were compatible, but represented different levels of the hierarchical structure of abilities. Even though Guilford (1981) accepted that his structure-of-intellect (S-O-I) model was defective, and reanalysed his data using oblique rotation, Brody (1992, p. 34) concluded that the purported factor structure underlying Guilford's model is seriously defective. In view of the lack of empirical support for Guilford's model, it does not provide a satisfactory alternative to the hierarchical Gf/Gc model.

Alternative structures also may be relevant. For example, Guttman's examination of the rank ordering of correlations suggested simplex, circumplex and radex structures. Simplex structures follow a linear sequence, whereas in circumplex structures, all variables lie on a circle, merging into each other. According to Romney and Bynner (1992), the simplex "is reflected in correlations that decrease from the principal diagonal of the correlation matrix to the corners; the 'circumplex' is shown by correlations that decrease initially and then increase towards the corners of the matrix. A 'radex' comprises circumplexes of tests of comparable complexity and simplexes of tests varying in complexity...." Radex theory, involving simplex and circumplex models, may be compatible with personality and intelligence structures.

Bynner and Romney (1986) argued that a split-simplex model is most appropriate, whereby vocabulary skill acts as a determinant of cognitive differentiation. This suggestion has also received support from cognitive information-processing research into memory (Schwartz & Reisberg, 1992). Brody (1992), and Marshalek, Lohman, and Snow (1983) showed that the factor analytically derived hierarchical model is compatible with Guttman's radex theory. Soldz, Budman, Demby, and Merry (1993) reported that whereas personality disorders can be meaningfully located in circumplex space, application of a hierarchical model enables more appropriate location of several disorders. Cattell (1983) and Eysenck (1991, 1992) have argued strongly for the importance of hierarchical models (see also the chapter by Stankov, Boyle, & Cattell). Clark, McEwen, Collard, and Hickok (1993, p. 90) reported on "the general utility of a dimensional approach to the assessment of personality disorder." According to John, Hampson, and Goldberg (1991), people prefer the highest level of abstraction in hierarchical trait models.

Popularity of hierarchical factor models reinforces the notion of stable traits, whereas simplex and circumplex models suggest that personality disorders are more responsive to therapeutic manipulation (Romney & Bynner, 1992). Disorders which can be modeled via circumplex theory may be more amenable to interpersonal psychotherapy, whereas those modeled by simplex theory might be managed best using cognitive-behavioral therapeutic techniques. Romney and Bynner (pp. 55-56) concluded that "parallelism between the circumplex and hierarchical factor models reflects the parallelism between the radex and hierarchical factor models...on abilities." However, Soldz et al. (1993) found that while many personality disorders could be located within the circumplex model, their placement within the hierarchical factor model provided a more accurate representation. In Zuckerman's (1991, p. xi) view, "the hierarchical model of traits...is best because it can encompass both broad and narrow traits. The alternate model of a circumplex is less useful because it is generally limited to a two-dimensional model."

Factor Analysis of Personality Traits

Several investigators (e.g., Cattell, 1983; Comrey, 1980; Eysenck, 1991; Guilford, 1975) have factor analyzed intercorrelations of personality variables with the aim of locating the major dimensions of human personality. This has resulted in the factor analytic development of several multidimensional instruments such as the 16PF (see Birkett-Cattell, 1989; Boyle, 1990b), the Comrey Personality Scales or CPS (Comrey, 1980), and the Eysenck Personality Questionnaire or EPQ (see Grayson, 1986). Zuckerman (1991, p. 54) alluded indirectly to one of the virtues of the 16PF, asserting that "a profile of scores on a multitrait test indicates which traits are salient...for a given individual without the need to devise an individualized idiodynamic assessment for every subject." The 16PF (and its junior versions--HSPQ and CPQ--see Schuerger, 1992) has stood the test of critical scrutiny over time in various editions of the *Mental Measurements Yearbooks* and/or *Test Critiques*. The 16PF measures 15 normal personality trait factors (plus Factor B--Intelligence) discerned factor analytically from examination of over 4000 trait names from the English dictionary. In addition, no fewer than six second-stratum factors have been discerned through factor analyses of the intercorrelations of the 16 scales. This multidimensional self-report instrument was constructed on the basis of a *comprehensive* assessment of the personality domain, as represented in the trait lexicon (cf. John, Angleitner, & Ostendorf, 1988). Moreover, *Cattellian psychology provides one of the few models which actively seeks to integrate the roles of personality and intelligence within the same psychometric instruments (e.g., 16PF/CAQ, HSPQ, CPQ)*--(see also chapter by Kline, on the critical assessment of measurement instruments).

Criticisms (e,g., Eysenck & Eysenck, 1985; Zuckerman, 1991) of attempts to replicate the 16PF primary factors based on item intercorrelations have not taken into account the unreliability of single-item responses. As Cattell (1973), Comrey (1980), and Marsh (1989) have all pointed out, it is essential to utilize more reliable groups of items (Cattell's item parcels; Comrey's FHIDs; Marsh's item-pairs). Mershon and Gorsuch (1988) have clearly demonstrated the importance of the

16PF primary factors in accounting for considerably more trait variance than do three or five factors.

Measures of psychopathological traits include the Minnesota Multiphasic Personality Inventory or MMPI and the MMPI-2 (Friedman, Webb, & Lewark, 1989), the Clinical Analysis Questionnaire or CAQ (Krug, 1980), and the Personality Assessment Inventory or PAI (Morey, 1991; see Boyle, 1993a). Eysenck (1991, p. 783) pointed out that nonfactorial models such as the MMPI, and California Psychological Inventory or CPI (Gough, 1987), inadequately measure personality structure. Eysenck (1985b) argued that it would make sense conceptually to factor analyze the CPI item intercorrelations (although because of item unreliability, an analysis of item parcels would be preferable). The resulting greater conceptual clarity would facilitate testing of psychological theories and models. Factor analytically derived scales are preferred over non-factored scales, especially in the clinical area, where extreme scores on factors may have etiological, diagnostic, and/or therapeutic implications. Moreover, according to Holden, Reddon, Jackson, and Helmes (1983, p. 37), "*factor analyses of the entire MMPI item pool...fail to support the original scoring keys* [italics added]." Helmes and Reddon (1993) provided an even more critical review of the MMPI and MMPI-2 instruments, pointing out that both instruments do not satisfy modern psychometric standards for assessing psychopathology. Since the factor structure of the MMPI does not seem consistent with its purported scale structure, its continued use can only serve to promulgate traditional psychiatric labelling and stereotyping. Hopefully, reliance on such archaic classifications will decline as we enter the 21st century.

Instruments such as the MMPI, CPI, or Hogan's (1986) Personality Inventory (HPI--see Boyle, 1992b) have been constructed using empirical-keying or an intuitive-rational approach, leaving doubt as to their scale validity. Soldz et al. (1993) provided evidence that psychopathology is often best viewed as the extremes of normally distributed traits, casting doubt on the validity of discrete diagnostic syndrome categories. The factor analytic evidence does not support the MMPI psychiatric syndrome structure (Holden et al., 1983; Eysenck, 1991). According to Eysenck (1991, p. 776), the MMPI includes

"ad hoc scales for arbitrarily chosen traits, without any personality theory in

> mind...when factor analysed the scales of the MMPI fail to appear as hypothesized, items correlate better with scales they do not belong to than with their proper scales...It is perhaps significant that the personality questionnaire more widely used than any other should violate all the rules laid down by psychometrists for the construction of such instruments; that it should be based on no recognizable or clearly stated theory of personality; and that the resulting scales should be interpreted in terms of highly subjective and scientifically meaningless categories."

Hence, these instruments, and other non-factor analytically verified instruments of their ilk, cannot be recommended for use in psychological assessment.

Simplification of the Multivariate Psychometric Model

A possible problem with the Cattellian psychometric model is that there are *too many primary factors* to be of practical utility for applied psychologists (at least 20 primary abilities, 16 normal personality traits, 12 abnormal personality traits, 8 emotional states, 10 motivation dynamic traits). Kline (1980, p. 324) pointed out that, "there are so many primaries that a workable useful model would involve so much testing time that it would not be viable. *If a model were to be used for any practical purpose, higher-order factors would have to be used* [italics added]." This preference for more parsimonious models of personality and ability structure has been emphasized also by John et al. (1991). One can reduce the number of dimensions by focusing on second-order factors (cf. Wiggins & Trapnell, in press).[2] Kline argued for incorporation of the *higher-stratum* factors from each intrapersonal psychological domain into a more parsimonious model. In this vein, Boyle has undertaken a programmatic series of studies into higher-order factors within the framework of the comprehensive Cattellian system, with the aim of producing a simplified, and more useful psychometric model. Boyle has delineated several second-stratum dimensions within each of the ability, personality trait, dynamic motivation, and mood-state spheres.

Across all intrapersonal psychological domains, the number of primary factors is considerable, whereas use of 25-30 second-stratum dimensions clearly enables greater ease of application. Yet, predictive validity is sacrificed in going from primary to secondary factors as shown by Mershon and Gorsuch (1988).

While 60-70 primary factors is a lot for busy clinicians to consider, (a) the truth of structure, and (b) the increasing quality of computer prediction should make it more acceptable to focus on primary factors (indeed, chemists deal with no fewer than 104 elements).

Higher-Stratum Cognitive Abilities

In accord with Cattell (1982b, 1987a), Horn and Stankov (1980), and Stankov and Horn (1982), higher-stratum abilities (Boyle, 1988b) have been labelled: fluid intelligence (Gf), crystallized intelligence (Gc), Memory Capacity (Gm), Perceptual Speed (Gps), Retrieval Capacity (Gr), Visualization Capacity (Gv), Auditory Organization (Ga), (cf. Kline & Cooper, 1984). Abilities can be viewed more easily in terms of this smaller number of secondaries. Gf and Gc offer an excellent example of the experimental verification of factors that Eysenck stresses. After being delineated as separate factors, it was found that (a) Gf has a sigma of 20 instead of Gc's 15 IQ points; (b) they follow totally different life courses; (c) Brain injury affects them differently; (d) they differ completely in suitability for cross-cultural comparison; and (e) Horn's results show that they differ independently as states from day to day. Cattell's *triadic theory of abilities* suggests that secondary ability factors comprise *general capacities* or *g*'s (fluid intelligence, Gf, crystallized intelligence, Gc, perceptual speed, Gps, retrieval capacity, Gr); *provincial powers* corresponding to the brain's visualization capacity, Gv, auditory organization, Ga, tactile and kinesthetic capacities; and *agencies* corresponding both to Gc and to Thurstone's primary ability factors (see Cattell, 1987; Woliver & Saeks, 1986).

Higher-Stratum Personality Dimensions

At least five higher-stratum personality dimensions have been verified within the normal trait domain (Boyle, 1989a). Previously, Cattell (1973, p. 116) had reported eight secondaries from over 10 separate factor analytic studies, (second-order in the 16PF personality sphere)--showing that Comrey's (1980) factors are closer to being true secondaries rather than primaries). Criticism of the use of factor analysis in delineating personality structure, cannot be justified on the superficial argument that Cattell, Eysenck, and Comrey have all claimed different numbers of factors. This criticism does not acknowledge that each investigator has focused on different levels within the hierarchical trait model. Within the abnormal domain, Boyle's (1987d) research has also suggested an additional six second-stratum dimensions, rather than the single P factor included in the EPQ. Higher-order psychopathological (CAQ) dimensions related to schizophrenia, psychopathy, psychotic inadequacy, paranoia, helpless depression, and anxious/agitated depression.

In the most comprehensive, methodologically sound scale factoring to-date of the 16PF on a sample of 17,381 subjects (crossvalidated for 9,222 males and 8,159 females separately), Krug and Johns (1986) confirmed *at least* five second-stratum dimensions in the normal trait domain, leaving little doubt as to their accuracy (they also extracted an intelligence factor, loading on Factor B). These findings also agree with Barton's Central State-Trait Kit or CST (see Cattell, 1973), in showing the importance of the secondary dimensions of Extraversion, Neuroticism/Anxiety, Conscientiousness, Tough Poise, and Control, and demonstrating the inadequacy of the plethora of less substantial claims, based on much smaller sample sizes (e.g., McKenzie, 1988; Mathews, 1989).

Five second-order personality factors have also been reported by several other investigators (see Digman, 1990; Goldberg, 1992). Claims as to their "robustness" (e.g., Costa & McCrae, 1992a) are misplaced, however (this term applies to departures from underlying statistical assumptions, such as multivariate normality, heteroscedasticity, etc.). It is a nonsense to talk about the "Robust Big Five." Doubts about the validity of the Big Five (Norman Five) have emerged (Livneh & Livneh, 1989), despite the claims of McCrae and Costa (1987, 1989), and Costa and McCrae (1992a,b,c). As Romney and Bynner (1992, p. 3) asserted, structuring the personality sphere into five broad dimensions, provides only one possible interpretation, which should not be regarded as irrefutable. Likewise, Eaves, Eysenck, and Martin (1989), and Eysenck (1990a,b, 1991), have concluded there is little empirical evidence to support the so-called Big Five (cf. John, 1990), proposing that only Eysenck's (1990a,b) Psychoticism (P), Extraversion (E) and Neuroticism (N) factors are needed. Whereas Eysenck's E and N factors emerge at the 16PF second-stratum level, Krug and Johns (1986) obtained an additional three normal dimensions at the Eysenckian level of analysis. In addition, Eysenck's P factor is represented by at least five abnormal trait dimensions at the CAQ second-stratum level (Boyle, 1987d)--(cf. Zuckerman, Kuhlman, & Camac, 1988).

Eysenck (1992) has criticized Costa and McCrae's (1992a) contention that the Big Five provide an adequate account of the normal personality sphere. According to Eysenck (p. 667), "the postulation of the 5-factor model is a premature crystallization of spurious orthodoxy." Eysenck suggested that apart from E and N, the remaining three dimensions proposed by Costa and McCrae are essentially primaries which are often intercorrelated highly. Eysenck pointed out that Costa and McCrae's work has ignored meta-analysis evidence which disputes their claim, and because they provide no theoretical underpinning or nomological network, or any attempt to relate the Big Five to underlying biological and neurobehavioral mechanisms (cf. Zuckerman, 1991, 1992). Eysenck (p. 668) concluded that "outside the narrow circle of 5-factor enthusiasts, research has completely failed to find basic factors similar to A, C or O." Thus, both Cattell and Eysenck are in complete agreement that studies of the so-called Big Five are scientifically unacceptable. Furthermore, as Clark, Vorhies, and McEwen (in press, pp. 33-34) pointed out, the five-factor model may not account for much of the emotion-related variance involved in maladaptive traits.

As Zuckerman (1991, p. 12) reported, "The theory of what was measured by the P scale...and the psychometric adequacy of the scale itself were challenged almost immediately...items in the scale are a mixture of impulsivity; sadism or lack of empathy; aggressiveness; sensation seeking; lack of concern about finances, work, or punctuality; uncommon social attitudes...and a few, mild paranoid-type items...items suggesting psychotic delusional thinking...were mostly dropped...because they were so infrequently endorsed..." Although Eysenck (1991) has summarized a number of factor analytic studies of the 16PF claiming support for his PEN system, none of these studies has been crossvalidated across thousands of subjects (as in the Krug & Johns, 16PF study). No statistical test (e.g., Bartlett, Scree, Humphreys-Montanelli, VSS, MAP etc.) will, on a personality sphere of variables, permit one to stop at five components in a first-order level factor analysis. One cannot obtain accurate secondaries by prematurely stopping a first-order extraction at the number of secondaries one "believes" exists! To do so only results in extraction and rotation of *pseudo-higher-order dimensions*. The gradual recognition of more factors is interesting: Spearman 1, Peabody 2, Eysenck 3, Thurstone 6, Comrey 8, Sells 11, and finally Cattell 16+.

The Big Five have emerged from suboptimal EFA procedures. The "Little Jiffy" approach generally produces inaccurate factor solutions, which nearly always fail to satisfy Thurstone's principle of simple structure (measured via the hyperplane count). This approach provides only a crude approximation to the actual factors (McDonald, 1985). The Big Five do not adequately measure the personality trait sphere, in contrast to Cattell's *comprehensive* measurement, covering at least five higher-stratum normal traits (16P/CAQ Part 1), a similar number of abnormal traits (CAQ Part 2), and 28 primary factors (16 normal; 12 abnormal). Cattell's (1987b) *Depth Psychometry* approach provides not only a *quantitative* assessment of personality, but also a *qualitative* account of each higher-stratum factor in terms of the loadings on particular primaries. Gough (1987) also tried to comprehensively measure the personality sphere (see also the chapter by Most & Zeidner), but his failure to utilize factor analytic methods in constructing the CPI resulted in an instrument with uncertain scale validities (Eysenck, 1991).

Also there has been undue restriction of trait variance, with only 20 of Cattell's original 36 trait clusters included in analyses resulting ultimately in the construction of instruments such as the NEO-PI (Costa & McCrae, 1992b,c). *Only 56% of the normal trait sphere is covered by the Big Five, so that claims as to their comprehensiveness are misplaced.* Indeed, Tupes and Christal (1961, p. 12) pointed out that "It is unlikely that the five factors identified are the *only* fundamental personality factors. There are quite likely other fundamental concepts involved among the Allport-Odbert adjectives..." Likewise, Norman (1963, p. 582) stated that "it is time to return to the total pool of trait names in the natural language...to search for additional personality indicators not easily subsumed under one or another of these five recurrent factors." In selecting only 20 of Cattell's original 36 personality rating scales, Norman biased his results in favor of the Tupes and Christal Big Five by selecting only those scales which were already known to be loaded most highly by them. Norman (p. 577) admitted that, "The four scales with the highest median factor loadings for each of the five factors identified in these earlier analyses were selected." Norman's study was prefaced on that of Tupes and Christal, who reported a partly inaccurate five-factor solution, derived from questionable factor analytic procedures (see Boyle, 1988c, for a discussion of factor analytic guidelines).

In contrast, Krug and Johns' second-order 16PF factors were based on (1) comprehensive sampling of the normal trait domain as measured in the 16PF (incorporating all 36 of Cattell's original clusters--not just 20); (2) use of appropriate factor analytic procedures on extremely large samples; (3) crossvalidation of findings on extremely large samples of males and females separately. Krug and Johns' factors correspond only approximately to the currently popular Big Five dimensions (Neuroticism, Extraversion, Openness, Agreeableness, Conscientiousness) which have an amended measurement basis in the NEO-PI (Costa & McCrae, 1992b). This lack of complete alignment also may be partly due to failure to achieve simple structure solutions. Boyle (1989a) showed that these currently popular dimensions correspond only roughly to the more reliable 16PF higher-stratum factors.

The five-factor solution presented by McCrae and Costa (1987) exhibited a $\pm.10$ hyperplane count of only 35.8%, indicating poor simple structure. Likewise, the solution presented by Costa and McCrae (1992a) exhibited poor simple structure (hyperplane count of only 30.3%). Costa and McCrae presented a five-factor solution for the Revised NEO-PI, which gave a hyperplane count of only 31.3%, again failing to achieve simple structure. In contrast, the factor solution reported by Krug and Johns (1986) gave a $\pm.10$ hyperplane count of 71.4%, suggesting much greater validity of the 16PF secondaries over the NEO-PI factors. Thus, the factor analytic basis of the Norman Five and the NEO-PI appears inadequate. According to Costa and McCrae, no fewer than six of the WAIS-R scales exhibit significant correlations with Factor O (Openness). The WAIS-R Block Design and Object Assembly subtests exhibit the largest correlations. Openness may be a hybrid dimension measuring both personality and intelligence variance in an unspecified way. Clark and Livesley (in press) also demonstrated that

Openness (Factor O) was neither strongly nor consistently verified.

Because of inadequate sampling of variables and subjects, and crude factor analytic procedures employed in construction of the NEO-PI, the resulting factors do not achieve a high level of simple structure. Costa and McCrae (1992a, p. 661) attempted to justify their weak factor solution, contending that "Although simple structure has been the guiding principle in factor analysis for decades, we know that personality traits do not necessarily conform to it...many important personality traits are defined by two or more factors..." Yet, empirical studies of physical plasmodes (see Cattell, 1978) have demonstrated that factors make sense only when simple structure emerges. Consequently, Costa and McCrae's assertion that simple structure does not matter, runs counter to accepted factor analytic principles (e.g., Cattell, 1978; Child, 1990; Gorsuch, 1983; McDonald, 1985).

Costa and McCrae (1991) argued that competing five-factor models of personality might be viewed as rotational variants of the NEO-PI Big Five. They reported an orthogonal re-rotation of the work of Zuckerman, Kuhlman, Thornquist, and Keirs (1991). However, most psychological constructs are correlated to some extent. If the aim of rotation is to achieve simple structure solutions, then oblique rotational procedures should be applied, systematically varying the degree of obliquity, and settling on that solution with the highest hyperplane count (cf. Cattell, 1978; Gorsuch, 1983). Costa and McCrae reanalysed the Zuckerman et al. data with the aim of supporting the claimed NEO-PI factor structure. However, perusal of their re-rotated solution indicates that simple structure was not achieved. Not only were there several instances where the same variables exhibited factor loadings in excess of .40 across the five factors. The hyperplane count was only 30.3%, suggesting the factor solution was not a simple structure one. Although Zuckerman et al. also used orthogonal rotation, their five-factor solution exhibited a hyperplane count of 38.2%. Costa and McCrae's attempt to verify their Big Five clearly failed as their factor solution was inferior to that provided by Zuckerman et al. (1991).

None of the studies into higher-stratum personality dimensions compare favorably with the Krug and Johns (1986) 16PF study, which was based on more comprehensive sampling of the normal trait domain, and much larger sample sizes (over 17,000 subjects as compared with just a few hundred subjects in the Costa & McCrae, and Zuckerman et al. studies). Evidently, research extending from the Norman Five to development of the NEO-PI appears deficient, both in sampling of subjects and variables, and in failure to attain simple structure solutions. As Zuckerman (1991, p. 17) pointed out, "The rallying around the "five robust factors," or the "big five" as their supporters call them, probably reflects disillusion with the...Cattell multifactor system and the feeling that Eysenck's big three are not enough dimensions to account for the complexity of personality."

Higher-Stratum Motivation and Mood-State Dimensions

Within the area of motivation dynamics, Boyle (e.g., 1985a, 1986b, Boyle, Start, & Hall, 1989) has delineated several higher-order dimensions as measured in the Motivation Analysis Test (MAT/SMAT/CMAT) series of instruments. These instruments are *objective* measures of dynamic motivation traits, and therefore essentially avoid the problems of response sets associated with item transparency. It is virtually impossible for respondents either consciously or unconsciously to distort their motivation profiles in any systematic way. Theoretically, much of human motivation is at the unconscious level (as measured by the unintegrated or U-components in the MAT/SMAT/CMAT instruments), so that comprehensive objective motivation measures are all the more important. Conscious motivation dynamic traits are measured via the (I) integrated components (see Cattell, 1985).

However, from these EFA studies, there is considerable discrepancy as to the specific nature of higher-stratum dimensions in studies of the MAT and School Motivation Analysis Test. For the 190-item SMAT (which measures drives labelled Assertiveness, Mating/Sex, Fear, Narcism, Pugnacity, Protectiveness; and acquired interest patterns labelled Self-Sentiment, Superego, School Orientation, and Home Orientation), Boyle et al. (1989) reported six higher-stratum factors among adolescents. Factor 1 contrasted U-Superego, U-School, and U-Home with U-Mating, and U-Narcism. Factor 2 loaded primarily on I-Pugnacity. Factor 3 loaded mostly on I-School and U-Pugnacity. Factor 4 contrasted I-Home and I-Protectiveness with (U + I) Pugnacity. Factor 5 loaded predominantly on U-School, U-Mating and I-Assertiveness, and finally Factor 6 contrasted I-Superego and (U +

I) Self-Sentiment with U-Fear. The 230-item Children's Motivation Analysis Test or CMAT (a downward extension of the MAT and SMAT instruments) provides measures of six biologically based *ergs* (labelled Narcism, Play, Fear, Pugnacity, Curiosity, and Assertiveness), and four culturally acquired *sentiments* (Home Orientation, Self-Sentiment, Superego, and School Orientation). In accord with the view that motivation dynamics are partly acquired as a function of development, higher-order analysis of the CMAT scale intercorrelations has revealed only four dimensions (e.g., Boyle & Start, 1989).

Since instruments designed to measure fluctuating states collectively cover some 30-40 primary mood-state dimensions, elucidation of a smaller set of central state factors would enable greater economy of measurement and administration time. Accordingly, Boyle (e.g., 1985a, 1987a, 1988a) carried out a series of programmatic studies into the higher-order factor structure of transitory mood states, using both single-occasion (R-factoring) and across-occasions (differential Dr-factoring) of change scores. Boyle (1987a) investigated the higher-stratum mood-state factors discernible from a conjoint Dr-factoring of difference scores for the 8SQ and DES-IV instruments. The net result has been elucidation of five to six central mood-state dimensions, which might be labelled *Extraversion State*, *Neuroticism State*, *Arousal-Fatigue*, *Hostility*, and *Curiosity*. Cattell (1979, 1980) has proposed an elaboration of his behavioral specification equation, incorporating trait-modulation indices to account for the influence of mood states on behavior. For each mood state within the Cattellian psychometric model, it is assumed that a state liability trait exits, on which individuals differ. A modulator expressing the mean stimulation of a given stimulus for a particular state transforms this liability value (i.e., situational indices modulate state liability traits).

Varieties of Psychometric Measurement Media

There are three different kinds of measurement media. The first is life-record (L-data) which includes *ratings* of others. L-data is often unreliable and invalid, as the perceptual and idiosyncratic biases of the rater may distort the picture of the individual being rated. Second is *questionnaire* (Q-data) which comprises an individual's self-ratings. Unfortunately, responses to transparent self-report questionnaire items are prone to distortion ranging all the way from inadequate self-insight to deliberate dissimulation (faking good or faking bad--see Bagozzi & Yi, 1990; Boyle, 1985b). Distortion may also occur due to the influence of response sets such as social desirability (Schmitt & Steyer, 1993). Third is *objective test* (T-data), wherein the items comprise non-transparent miniature performance tests (see Cattell & Warburton, 1967, for a compendium of over 500 such miniature objective test devices). T-data has the advantage that response distortion is minimized, as there is no immediately discernible relationship between item content and corresponding personality or ability factors being measured (Boyle, 1990c; Schuerger, 1986; Schmidt, 1988).

Personality inventories such as the 16PF/CAQ, HSPQ, CPQ, Myers-Briggs Type Indicator or MBTI (Briggs-Myers & Briggs, 1985), CPI, HPI, NEO-PI, MMPI and MMPI-2, CPS (Comrey, 1980), and EPQ (Eysenck, 1991), have all utilized self-report Q-data. However, instruments for measuring intellectual abilities have been based on objective T-data from the very beginning, starting with Galton's simple RT studies (Jensen, 1991), extending to the SB-IV and the Wechsler Intelligence Scales (WAIS-R, WISC-R, WPPSI)--(cf. Kaufman, 1990). Factor analytic work has resulted in development of the 16PF, CAQ, MAT, CAB, and O-A Batteries (see Cattell & Johnson, 1986, for a detailed description). Such multidimensional instruments are "hostages" for the ability and personality structure theory, enabling *empirical measurement* of factors (mapped longitudinally as life course curves), and quantification of heritability estimates (via Multiple Abstract Variance Analysis or MAVA--see Cattell, 1982b, pp. 89-123), along with *experimental investigation* of abilities and personality traits.[3] The MAVA method provides a sophisticated analysis of the contributions of genetic and environmental variance. As Kline (1993a, p. 102) reported, "Jinks and Fulker (1970) indeed describe it as a brilliant one-man attempt to develop a statistics of genetic biometrics..."

Motivation/Response Distortion in Personality Questionnaires

It is possible to modify Q-data for response bias using *motivational distortion scales* which are built into the 16PF range of instruments (Birkett-Cattell, 1989). However, on the basis of trait-view theory (Cattell, 1982a, 1992b), such modification of scale scores is potentially problematic, and reliance on traditional motivation distortion scales may only serve to add further measurement noise into responses on psychometric instruments. Holden, Kroner, Fekken, and Popham (1992) have shown that when faking good, individuals take relatively longer to respond to socially undesirable Q-data items, and *vice versa*. Holden et al. (p. 272) stated that, "The model predicts that differential test item response latencies should be faster for schema-congruent test answers than for noncongruent responses." Since virtually all personality instruments utilize Q-data, whereas intelligence tests are based on T-data measures, the measurement of personality traits has not yet reached the level of certainty already achieved with intelligence testing (Brody, 1992). Failure to obtain simple structure solutions in factor analytically constructed instruments has only served to confound research findings. Thus, Guilford's S-O-I model, the MMPI, MBTI, HPI, CPI all fail to satisfy simple structure requirements. The scale structures of these instruments are not supported factor analytically.

Need for Objective Personality Test Construction

Intelligence tests are based on performance T-data, whereas almost all personality instruments (e.g., CPI, MMPI, CPS, EPQ, 16PF, HSPQ, CAQ, NEO-PI, 8SQ, DES-IV, POMS, MDQ) are merely self-rating Q-data scales. There is an urgent need for construction of *multivariate objective personality tests*, along the lines of the Objective-Analytic (O-A) Battery (Cattell & Schuerger, 1978) which measures 10 factor analytically derived personality traits (see Gough, 1989). Objective (T-data) tests avoid self-report distortion and rater bias. In an objective personality test, the respondent does not know which particular trait is being measured. Scales measured in the O-A Battery have been labelled (using *Universal Index Numbers*): U.I 16 Ego Standards; U.I. 19 Independence vs. Subduedness; U.I. 20 Evasiveness; U.I. 21 Exuberance; U.I. 23 Capacity to Mobilize vs. Regression; U.I 24 Anxiety; U.I. 25 Realism; U.I. 28 Asthenia vs. Self-Assurance; U.I. 32 Exvia vs. Invia; U.I. 33 Discouragement vs. Sanguiness. These T-data factors correspond to the second-order 16PF factors, raising questions about the meaningfulness of the primary Q-data trait dimensions. The O-A primary factors correspond to normal and abnormal personality L- and Q-data traits at the second-stratum level. Each factor is measured on seven or eight subtests, taking 20-30 minutes each. According to Bolton (1988), the O-A Battery represents an innovative approach to personality assessment. Given the realities of testing in practical settings, administration of the complete O-A Battery is likely to take appreciably longer than the nominal five hours. Some evidence of the predictive validity of the O-A Battery in discriminating between various psychiatric syndromes has been provided (see Schuerger, 1986).

Objective Motivation Measurement

Within the dynamic motivational sphere--measured via the MAT/SMAT/CMAT series of instruments--application of the behavioral specification equation yields interesting new indications of total motivation (U + I scores), conflict (U-I scores), derivative scores (e.g., Information-Intelligence), decision theory, interests, and so on, not yet experimentally investigated. The role of transitory states is also included, as researched in several studies by Boyle (e.g., 1983a, 1985a, 1987a,b, 1988a). There are various new concepts, as models here. For example, vector measurement of interest and learning in the dynamic lattice; matrix calculation of learning in life selections; multiple factoring in the data box combining factors of persons x stimuli x occasions; the vector representation of environment; modulation law of states; assignment of vulnerability indices to tests; factor analytic discovery of states by dR- and P-techniques; the representation of learning gain by vector change; law of structured learning through gain in dynamic structures; and representation and quantification of perception change (construing a contextual emphasis in trait view theory: "attribution theory"), and so on (see Cattell, 1979, 1980).

Item Analysis Issues: Psychometric Properties in Personality and Intelligence Research

Reliability: Stability vs. Dependability

Reliability of psychometric scales is an important precursor for validity (see Thorndike, 1982). It is consistency as measured over time (test-retest rather than "internal consistency") which provides the most accurate estimation of reliability (cf. Fernandez, 1990; Fernandez, Nygren, & Thorn, 1991). It is important to differentiate between short-term *dependability* (immediate test-retest) versus longer-term *stability* (retest intervals ranging from say one week to several years) which allows an estimation of measurement error (Cattell, 1973). This distinction is critical in assessing the reliability of state versus trait instruments (Boyle, 1983a). Both dependability and stability estimates should be high (.8 or .9) for trait measures (e.g., CPI, MMPI, HPI, MBTI, EPQ, 16PF, CAQ, etc.). For state measures, however, dependability estimates should be high, while stabilities should be considerably lower if the scale is truly sensitive to situational variability (Boyle, 1985b). For example, the State-Trait Anxiety Inventory or STAI (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1986) should exhibit high dependability estimates for both sections of the instrument, but stability coefficients should be appreciably higher for trait as compared with state scales. In regard to the O-A Battery, Bolton (p. 378) reported test-retest reliabilities over a one day interval ranging from .62 to .93 (median .75), and stability coefficients for retest over three to six weeks ranging from .61 to .85 (median .71). Concept validities (correlations between scale scores and the pure factors) ranged from .64 to .92 (median .76). Consequently, some of the T-data factors measured in the O-A Battery are less stable than desirable for measures of enduring personality traits. In comparing the reliability and validity of personality and ability instruments, intelligence tests (T-data measures) exhibit appreciably higher coefficients in both respects than do most personality inventories.

Item Homogeneity: Internal Consistency vs. Item Redundancy

Reliability is a function of the length of a scale in accord with the Spearman-Brown prophecy formula (Crocker & Algina, 1986). In general, longer scales with a larger number of items are more reliable than are shorter scales. The item homogeneity of a scale should not be excessively high, otherwise "internal consistency" may become "item redundancy," whereby items are virtually paraphrases of each other (Boyle, 1991a). We challenge the commonly held view that item homogeneity should always be maximized. Indeed, Cattell (1978) has indicated that low to moderate item homogeneity is preferable, so that each item contributes to the breadth of measurement of a particular scale. Kline (1986) suggested that item homogeneities in the 0.3 to 0.7 range are most desirable. According to Kline (1986, pp. 2-3), "Cattell argues that high internal consistency is actually antithetical to validity on the grounds that any item must cover less ground or be narrower than the criterion we are trying to measure...This is obviously the case, for if two variables were perfectly correlated, one would be providing no new information. Thus maximum validity, in Cattell's argument, is obtained where test items do not all correlate with each other, but where each correlates positively with the criterion. Such a test would have only low internal-consistency reliability. In my view, Cattell is theoretically correct." (cf. Cattell, 1982c).

Item Response Theory and Computerized Adaptive Testing

Item response theory (IRT)--sometimes termed latent trait theory--has emerged as a result of shortcomings with classical test theory or CTT (see Hambleton & Swaminathan, 1985, pp. 1-4, for a list of limitations). IRT relates item responses to an underlying ability trait, where the probability of a correct response to a given item is a function of the ability level. The probability of a correct response takes the graphical shape of an ogive, which can be defined by up to three parameters (item difficulty--position of curve relative to X-axis; item discrimination--slope of curve; and "guessing" parameter--lower asymptote of curve). If these parameters are known, it is necessary to estimate individuals' ability in terms of their item responses; to select the next item to be presented; and to constantly update their ability estimates. Use of CAT enables estimation of an individual's trait from as few as half the usual number of administered items. Individuals receive different numbers and combinations of items depending on their particular responses to items (i.e., their ability levels). CAT is considerably more efficient than standard tests constructed on the basis of classical test theory (cf. Crocker & Algina, 1986; Wainer, Dorans, Flauger, Green, Mislevy, Steinberg, & Thissen, 1990). A number of major aptitude and ability instruments have been constructed using IRT methods (e.g., US Armed Services Vocational Aptitude Battery or CAT-ASVAB; DAT).

With recent developments in IRT it is possible to check the contribution of individual items to total scale scores, enabling decisions as to which items to retain, and which to remove. Some items may exhibit significant measurement error. The issue of response bias in relation to scoring formats across items with differing levels of measurement also can be addressed more effectively with IRT than with CTT methods (which include the Spearman-Brown prophecy formula, standard error of measurement, Kuder-Richardson estimates of item homogeneity, and dissatenuation statistics). As pointed out by Hambleton, Swaminathan, and

Rogers (1991), CTT measures of item homogeneity or internal consistency (such as Cronbach's alpha coefficient) do not test the adequacy of summed scale scores (cf. Boyle, 1991a). In contrast to CTT methods, IRT is associated with a number of psychometric advances including the facility to evaluate items for their bias, difficulty level, and relationships to other items within a scale (Rudy, Turk, & Brody, 1992).

Use of IRT allows scale-free measures to be developed, so that various sets of items with scaling and measurement properties can be incorporated into equivalent versions of a scale. Since item parameters associated with IRT approaches are theoretically sample independent, item banks can be readily formed. Unlike traditional CTT models, IRT models are potentially falsifiable, and the statistical fit of specific items and the total scale score are tested explicitly. IRT methods also handle missing item data well, and enable tests of the legitimacy of "estimated scales" (Rudy et al., 1992).

It is desirable to employ both CTT and IRT methods, where possible, in the construction of intelligence and psychological tests and instruments. On the assumption of a unidmensional model (ascertained by means of an initial factor analysis), one- two- and three-parameter IRT models can be employed. The simplest is the one-parameter or Rasch model which differentiates between items solely in terms of their "difficulty" levels. However, the practical utility of the Rasch model has been severely criticised by Goldstein (1980). Certain considerations must be taken into account in deciding which model is the appropriate one (see Hambleton & Swaminathan, pp. 307-308, for a discussion of these issues). An interesting application of the two-parameter logistic model has been the work of Grayson (1986), who investigated latent trait models of dichotomous personality questionnaire (EPQ) data.

More advanced IRT models (e.g., partial-credit models) overcome some of the problems associated with Rasch scaling, but these in turn, introduce new variables and therefore the possibility of additional error (Hutchinson, 1991). Use of IRT provides evidence as to the most efficient items in a scale (cf. Butcher, Keller, & Bacon, 1985, regarding extending adaptive testing to personality instruments). The application of computerized adaptive testing (CAT), and more advanced IRT methods in personality and intelligence assessment, is likely to increase dramatically as we enter the 21st century. As Hambleton and Swaminathan have pointed out, IRT methods facilitate the equating of test scores, item bank development (sets of items with equivalent item characteristics), detection of biased items, and resulting psychological test construction. In the future, greater emphasis will be placed on developing multidimensional IRT models (Weiss & Yoes, 1990), which will have many implications for the construction of personality and intelligence tests.

Correlation Coefficients with Ordinal or Categorical Data

Another potential problem at the item analysis level is the computation of Pearson product-moment correlation coefficients when variables are ordinal or categorical. These estimates can be significantly biased as demonstrated empirically by Jöreskog and Sörbom (1988). It is desirable therefore, to compute polychoric correlations when the data are categorical or ordinal (cf. Hambleton & Swaminathan, 1985), and polyserial coefficients when ordinal and continuous variables are correlated, to minimize biased estimates (Poon & Lee, 1987). Use of PRELIS (Jöreskog & Sörbom, 1988) enables the simultaneous computation of polychoric, polyserial, and product-moment correlation coefficients as required, depending on the measurement level of each pair of variables being correlated. Since the computation of such correlation estimates is the starting point for many multivariate statistical procedures, it is essential that the best possible correlation estimates be derived in the first instance (Boyle, 1991c). Some of the major statistical packages provide estimates only of the product-moment correlations, thereby reducing the validity of many statistical analyses. Consequently, the resultant measurement error built into computed correlation coefficients will be compounded at every subsequent step in the data analysis procedures.

Statistical Effect Size

One of the difficulties with quantitative analyses of data is the distinction between statistical and practical significance. Although treatment effects may be statistically significant, often these effects are trivial and of little practical or conceptual meaningfulness. This issue is particularly problematic when dealing with multivariate analyses based on data from large samples (many personality and intelligence tests, for example, are multidimensional in structure), as the probability of obtaining statistically significant, but trivial effects, is increased in direct proportion to the number of scales. For example, the CPI comprises 20 trait scales, the CAQ has 28 scales, the PAI includes 22 scales, the CAB measures 20 primary abilities, and the NEO-PI-R includes no fewer than 30 primary scales!

Application of the Bonferroni correction reduces the likelihood of accepting statistically significant but trivial results (Winer, Brown, & Michels, 1991). Another approach is to calculate the corresponding effect sizes for each significant effect. Calculation of effect sizes rather than merely relying on simple significance test results *per se*, builds into the analysis the requisite degree of caution necessary to draw useful conclusions regarding the size of treatment effects. Interpretation of multivariate analyses of multidimensional personality and intelligence measures requires careful consideration of whether or not significant statistical effects have any practical or conceptual meaning.

Generalizability Procedures

Generalizability theory (Cronbach, 1990; de Gruijter & van der Kamp, 1990; Shavelson, Webb, & Rowley, 1989) involves a generalization of classical test theory. Whereas true-score theory assumes that error variance is homogeneous and that there is only one true score, generalizability theory differentiates between sources of error, enabling a quantitative estimation of the various error components. Construct interpretation is thereby facilitated by a knowledge of which sources of error are larger than others. This a critically important issue, as the reliability of personality and intelligence measures must be viewed in the light of likely error rates. Three salient sources of error include situations, occasions of measurement, and actual observations. As a general measurement procedure, generalizability theory involves observation, estimation, measurement, and optimization stages. Unfortunately, application of multivariate generalizability has not received much attention in the literature to-date (de Gruijter & van der Kamp, 1990; Webb, Shavelson, & Maddahian, 1983). However, an interesting application to personality research was undertaken by van Heck (1988), wherein the generalizability of L-data and Q-data across situations was investigated.

We have already discussed many of the likely sources of error associated with L-data and Q-data (see section above on the psychometric measurement of abilities and personality traits). However, both the situation and the occasion of measurement are subject to situational variability and fluctuation of mood states. Cattell (1979) distinguished between the ambient situation (*k*) versus the overall global situation (*e*), as likely sources of error, and built these sources of variability into the more complex versions of his behavioral specification equation, proposing the notion of modulation theory and state-liability traits). A simplified representation (where the global situation is assumed to comprise the focal stimulus, *h*, plus the ambient situation) is:

$$a_{hijk} = b_{hj1} \mathrm{T}_{1i} + \ldots + b_{hjp} \mathrm{T}\boldsymbol{pi} + \ldots + b_{hjs1} s_{k1} \mathrm{L}_{1i} + \ldots + b_{hjsq} s_{kq} \mathrm{L}_{qi} + \text{uniqueness} \quad \text{------------}(6)$$

where the modulator index is s_{kx} for trait x, i represents the individual, j is the response, s denotes the ambient situation indices, and L represents the individual's liabilities (see Cattell, pp. 187-196). Fortunately, application of the newer SEM methods via statistical modeling packages such as LISREL, COSAN, and EQS can facilitate estimation of error terms, thereby enhancing the importance of generalizability theory in personality and intelligence research.

Generalizability theory as extended into the multivariate context, is related to covariance structure analysis. Both approaches attempt to obtain estimates of variance from the variance-covariance matrix. However, the underlying assumptions for generalizability studies are generally weaker than those associated with covariance structure analyses (Brennan, 1983; de Gruijter & van der Kamp, 1990).

Test Bias in Personality and Intelligence Research

The purpose of administering intelligence and/or personality tests is to make valid predictions of future behaviors. Culture-fair tests (such as Cattell's CFIT measures discussed above) go part of the way in facilitating accurate predictions across different societies. However, by restricting the content of these instruments to that which is common across cultures, the relationship of this content to real-life situations, the predictive validity may be lowered inadvertently. Consequently, many personality and intelligence tests are significantly biased against one cultural group or another. Use of appropriate norms which pertain to particular groups or subgroups, is important if biased interpretation of test scores is to be avoided. As Anastasi (1990, p. 194) pointed out, "Validity coefficients, regression weights, and cut-off scores may vary as a function of differences in the test takers' experiential backgrounds." Hunter, Schmidt, and Rauschenberger (1984) provided an extensive review of cultural and ethnic effects on predictive validity of standardized psychological test scores. Possible test bias (both slope bias and intercept bias--see Anastasi, pp. 194-199) has been suggested, particularly in the area of intelligence testing. It has been reported that different cultural groups perform differentially on standardized scales. Nevertheless, according to Anastasi (p. 197), in the USA, "comprehensive surveys and critical analyses...have failed to support the hypothesis that ability tests are less valid for blacks than for whites in predicting occupational or educational performance..." Furthermore, there is no clearcut evidence of intercept bias, even when a test exhibits similar validity across cultural or ethnic subgroups (Anastasi, p. 199; Hunter et al., 1984).

Summary and Conclusions

The scientific analysis of personality and intelligence now predominates over the earlier, more subjective philosophical and literary speculations. Classical bivariate experimental designs in psychological research have been unduly emphasized at the expense of more appropriate multivariate experimental designs. In contrast to this univariate approach, the Cattellian school stands out as a major force in the promotion of multivariate experimental methods in both the ability and personality domains. However, the Cattellian psychometric model incorporates many primary abilities, personality traits, dynamic motivation factors, and transitory mood-state dimensions. While primary traits measured in instruments such as the 16PF and CAQ are numerous, second-order dimensions are more reliable, due to the greater number of items loaded by each secondary. Yet, higher-stratum dimensions are less predictive than primary traits. However, in line with Cattell's *Depth Psychometry* (Cattell, 1987b), 16PF second-order factors can be interpreted qualitatively in terms of their unique loadings on each of the contributing primary trait factors.

Krug and Johns (1986) demonstrated *at least* five major normal personality dimensions, in addition to intelligence. Other investigations (exemplified in the NEO-PI) suggest a slightly different breakdown of the personality sphere. However, *the so-called "Big Five" have a history plagued by inadequate sampling of subjects, variables, and inadequate EFA procedures).* Norman's (1963) study (on which the Big Five is prefaced) was flawed in (1) accounting for just over half the known personality trait variance; and (2) in its use of an inappropriate orthogonal rotation which precluded the possibility of obtaining a simple structure solution. Norman's five-factor solution closely matched that of Tupes and Christal, since the variables were highly selected to maximize the likelihood of finding the Big Five. This approach to research is to be abhorred. The currently popular Big Five provide an inadequate overview of personality trait structure.

Krug and Johns (1986) more comprehensive factor solution satisfied simple structure requirements (hyperplane count of 71.4%). In contrast, McCrae and Costa's (1987) factor solution for the NEO-PI exhibited a hyperplane count of only 35.8%, and the corresponding hyperplane count for the Zuckerman et al. (1991) study was only 38.2%, raising doubts about simple structure. At the second-stratum level, *at least* five major abnormal trait dimensions also emerge from factor analyses of the CAQ primary trait intercorrelations (Boyle, 1987d).

Although the Cattellian and Eysenckian schools appear to differ with respect to the number of personality trait dimensions, this is more a question of interpretation, than an insurmountable barrier. Both Cattell and Eysenck agree on the importance of factor analysis in psychometric research, and both agree on many substantive issues (thus the second-order 16PF dimensions correspond closely with the Eysenckian factors), as Eysenck (1984) has indicated. In reviewing the commonality between the Cattellian and Eysenckian schools, Eysenck (p. 336) stated that "the major conclusions are surprisingly alike; the only remaining difference is that Cattell attaches more importance than I do to his primary factors...it is unusual to discover such close correspondence between authors so distinct in their methods, procedures, evaluations and premises...The Cattell and Eysenck constructs and theories should be seen, not as mutually contradictory, but as complementary and mutually supportive." Moreover, as Boyle (1989, pp. 1296-1297) has pointed out, "Arguments against the importance of hierarchical structural models of personality, and against the use of factor analysis in discovering and confirming personality structure, cannot be justified on the superficial assertion that Eysenck, Comrey and Cattell have proposed different numbers of trait dimensions. This frivolous argument fails to acknowledge that each investigator has focused his attention on different levels within the hierarchical structural model of personality traits." The focus of Boyle's work has been on development of a more parsimonious version of the Cattellian psychometric model, emphasizing second-stratum factors instead of primaries. This not only enhances the practical utility of the model, in line with the Eysenckian emphasis on typological dimensions, but extends greatly the coverage of each of the major intrapersonal psychological domains of abilities, traits, dynamics, and states at the broad Eysenckian level of analysis.

Future psychometric research should also focus on construction of objective personality (T-data) instruments to minimize problems associated with item-transparency, response bias and motivational distortion. Although to-date, most effort has concentrated on personality inventories, these instruments are highly susceptible to motivation distortion. One objective personality instrument (the O-A Battery) has not received widespread use, partly because of the excessive administration time (at least five hours), so that its utility in applied situations has not yet been fully explored. Although the O-A Battery enables objective measurement of personality traits, it has not yet received sufficient usage to clearly assess its psychometric properties. Research into *objective* (T-data) measures of personality undoubtedly offers much promise for a more scientific approach to personality assessment, taking into account underlying psychobiological mechanisms (cf. Zuckerman, 1991).

The two-handed approach of EFA followed by CFA on an independent sample is a logical way to proceed. The two approaches serve entirely different purposes, and are complementary rather than competing methods. A frequent criticism is that EFA yields unstable factors, which seldom agree with the results of other investigators. An example is the apparent discrepancy between the Eysenckian and Cattellian personality factors, Eysenck claims three major trait dimensions, whereas Cattell examines 16 primary factors, and six secondary dimensions in the normal trait domain alone (clearly to cover *both* the normal and abnormal personality trait domains, *at least* 10-12 higher-stratum dimensions--"Big Ten" or "Big Twelve"--are required). It is argued that this difference in number of factors demonstrates the unreliability of EFA methods. However, this criticism is invalid. Second-order factoring of the Cattellian primaries produces the Eysenckian factors together with several additional trait dimensions at the Eysenckian level of analysis (*the Eysenckian model of personality structure accounts only for about 25-30% of the variance measured within the comprehensive Cattellian framework; likewise, Comrey's system accounts for only 67-80% of the Cattellian variance*). The Cattellian and Eysenckian factor analytic results exhibit much convergence, as long as comparisons are made at the appropriate level of the hierarchical structure of personality (i.e. at the Cattellian second-stratum level).

Provided adequate sampling of subjects and variables and appropriate methods of factor analysis are employed (cf. Cattell, 1978; Gorsuch, 1983; McDonald, 1985), and simple structure is obtained (and verified), EFA is an invaluable tool for mapping out the dimensionality of a domain (CFA can then be used to test the validity of the proposed factors). *Even with sample sizes of 300 subjects, the correct (exploratory) factor pattern solution is obtained only 50% of the time*. This demonstrates the importance of utilizing large samples when undertaking analyses of multivariate data. Many of the published EFA studies have been defective on various methodological grounds. Often studies have not paid adequate attention to this crucial issue, leading to the false impression that EFA is unreliable because it is "sample-driven" whereas CFA is more reliable because it is "conceptually-driven." It is nonsense to assume that confirmatory methods are not influenced by the idiosyncrasies of the samples.

Aside from the new possibilities for research into personality and intelligence using multilevel modeling packages, the general advantage of SEM models over the older exploratory methods (factor analysis, multiple regression analysis, path analysis) is that psychological dimensions can be modeled *dynamically*, wherein change in one aspect might be viewed as "causing" changes in another. Thus, in regard to cognitive abilities, vocabulary appears to directly affect both verbal, and numerical abilities (vocabulary is a precursor for growth of cognitive skills). According to Romney and Bynner (1992, p. 100), "personality disorders are more complicated, half...lying on a straight line (simplex) and half lying on a circle (circumplex)...Personal characteristics, whether they be intellectual, behavioral, or attitudinal, are all amenable to change." Soldz et al. (1993), however, demonstrated that the hierarchical factor model of personality structure is better able to account for personality disorders than is the circumplex model. Therefore, it is important to re-evaluate the adequacy of the so-called Big Five, and to appreciate the need for a more comprehensive coverage of the trait sphere than that provided in instruments such as the NEO-PI.

Personality and intelligence tests play a complementary role in the assessment of psychological functioning. However, many of the extant instruments have severe psychometric limitations pertaining to their psychometric properties, including basic reliability and validity. Nevertheless, with the advent of modern CFA, SEM and multilevel modeling techniques, we can now confidently expect some major advances in the psychometric conceptualization, measurement and statistical models of personality and intelligence.

References

Anastasi, A. (1990). Psychological testing (6th ed.). New York: Macmillan.

Anderson, J. G., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin, 103, 411-423.

Bagozzi, R. P., & Yi, Y. (1990). Assessing method variance in multitrait-multimethod matrices: The case of self-reported affect and perceptions at work. Journal of Applied Psychology, 75, 547-560.

Barrett, P. T., & Kline, P. (1982). Factor extraction: An examination of three methods. Personality Study and Group Behavior, 2, 84-98.

Bentler, P. M. (1988). Bentler critiques structural equation modeling. The Score, October, 3 & 6.

Bentler, P. M. (1989). EQS: Structural equations program manual. Los Angeles: BMDP Statistical Software.

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107, 238-246.

Bentler, P. M., & Kano, Y. (1990). On the equivalence of factors and components. Multivariate Behavioral Research, 25, 67-74.

Biddle, B. J., & Marlin, M. M. (1987). Causality, confirmation, credulity, and structural equation modeling. Child Development, 58, 4-17.

Birkett-Cattell, H. (1989). The 16PF: Personality in depth. Champaign, IL: Institute for Personality and Ability Testing.

Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

Bolton, B. (1988). Review of the Objective-Analytic Batteries. Test Critiques, 7, 374-380.

Boyle, G. J. (1983a). Critical review of state-trait curiosity test development. Motivation and Emotion, 7, 377-397.

Boyle, G. J. (1983b). Effects on academic learning of manipulating emotional states and motivational dynamics. British Journal of Educational Psychology, 53, 347-357.

Boyle, G. J. (1985a). A reanalysis of the higher-order factor structure of the Motivation Analysis Test and the Eight State Questionnaire. Personality and Individual Differences, 6, 367-374.

Boyle, G. J. (1985b). Self-report measures of depression: Some psychometric considerations. British Journal of Clinical Psychology, 24, 45-59.
Boyle, G. J. (1986a). Clinical neuropsychological assessment: Abbreviating the Halstead Category Test of brain dysfunction. Journal of Clinical Psychology, 42, 615-625.

Boyle, G. J. (1986b). Intermodality superfactors in the Sixteen Personality Factor Questionnaire, Eight State Battery and objective Motivation Analysis Test. Personality and Individual Differences, 7, 583-586.
Boyle, G. J. (1987a). A conjoint dR-factoring of the 8SQ/DES-IV multivariate mood-state scales. Australian Journal of Psychology, 39, 79-87.
Boyle, G. J. (1987b). Central dynamic traits measured in the School Motivation Analysis Test. Multivariate Experimental Clinical Research, 9, 11-26.
Boyle, G. J. (1987c). Commentary: The role of intrapersonal psychological variables in academic school learning. Journal of School Psychology, 25, 389-392.
Boyle, G. J. (1987d). Psychopathological depression superfactors in the Clinical Analysis Questionnaire. Personality and Individual Differences, 8, 609-614.
Boyle, G. J. (1987e). Use of change scores in redundancy analyses of multivariate psychological inventories. Personality and Individual Differences, 8, 845-854.
Boyle, G. J. (1988a). Central clinical states: An examination of the Profile of Mood States and the Eight State Questionnaire. Journal of Psychopathology and Behavioral Assessment, 10, 205-215.
Boyle, G. J. (1988b). Contribution of Cattellian psychometrics to the elucidation of human intellectual structure. Multivariate Experimental Clinical Research, 8, 267-273.
Boyle, G. J. (1988c). Elucidation of motivation structure by dynamic calculus. In J. R. Nesselroade & R. B. Cattell (Eds.), Handbook of multivariate experimental psychology (2nd ed.),
(pp. 737-787). New York: Plenum.
Boyle, G. J. (1989a). Re-examination of the major personality-type factors in the Cattell, Comrey, and Eysenck scales: Were the factor solutions by Noller et al. optimal? Personality and Individual Differences, 10, 1289-1299.
Boyle, G. J (1989b). Reliability and validity of the Stanford-Binet Intelligence Scale

(Fourth Edition) in the Australian context: A review. Australian Educational and Developmental Psychologist, 6, 21-23.

Boyle, G. J. (1990a). Stanford-Binet Intelligence Scale: Is its structure supported by LISREL congeneric factor analyses? Personality and Individual Differences, 11, 1175-1181.

Boyle, G. J. (1990b). A review of the factor structure of the Sixteen Personality Factor Questionnaire and the Clinical Analysis Questionnaire. Psychological Test Bulletin, 3, 40-45.

Boyle, G. J. (1990c). Integration of personality and intelligence measurement within the Cattellian psychometric model. Paper presented at the symposium on personality and intelligence, Fifth European Conference on Personality, University of Rome, Italy, June 12-15.

Boyle, G. J. (1991a). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? Personality and Individual Differences, 12, 291-294.

Boyle, G. J. (1991c). Experimental psychology does require a multivariate perspective. Contemporary Psychology, 36, 350-351.

Boyle, G. J. (1991d). Item analysis of the subscales in the Eight State Questionnaire (8SQ): Exploratory and confirmatory factor analyses. Multivariate Experimental Clinical Research, 10, 37-65.

Boyle, G. J. (1992a). Factor structure of the Menstrual Distress Questionnaire (MDQ): Exploratory and LISREL analyses. Personality and Individual Differences, 13, 1-15.

Boyle, G. J. (1992b). Hogan Personality Inventory. Psychological Test Bulletin, 5, 130-138.

Boyle, G. J. (1993a). Review of the Personality Assessment Inventory (PAI). In J. C. Conoley & J. Impara (Eds), Twelfth Mental Measurements Yearbook. Lincoln, NE: Buros Institute of Mental Measurements.

Boyle, G. J. (1993b). Relationships between mood states and menstrual cycle symptoms in young, healthy women: A structural modeling analysis. Unpublished manuscript.

Boyle, G. J. (1993c) Review of the Self-Description Questionnaire (SDQ-II). Test Critiques, 10.

Boyle, G. J. (1993d). Special review: Evaluation of the exploratory factor analysis programs provided in SPSSX and SPSS/PC+. Multivariate Experimental Clinical Research, 10, 129-135.

Boyle, G. J. (in press). Intelligence and personality measurement within the Cattellian psychometric model. In G. L. Van Heck, P. Borkenau, I. Deary, & W. Nowack (Eds.), Personality psychology in Europe, Vol. 4.

Boyle, G. J., & Cattell, R. B. (1984). Proof of situational sensitivity of mood states and dynamic traits--ergs and sentiments--to disturbing stimuli. Personality and Individual Differences, 5, 541-548.

Boyle, G. J., & Fabris, S. (1992). LISREL analyses of the RIASEC model: Confirmatory and congeneric factor analyses of Holland's Self-Directed Search. Personality and Individual Differences, 13, 1077-1084.

Boyle, G. J., & Langley, P. D. (1989). Elementary statistical methods: For students of psychology, education and the social sciences. Sydney: Pergamon.

Boyle, G. J., & Stanley, G. V. (1986). Application of factor analysis in psychological research: Improvement of simple structure by computer-assisted graphic oblique transformation: A brief note. Multivariate Experimental Clinical Research, 8, 175-182.

Boyle, G. J., & Start, K. B. (1989). Comparison of higher-stratum motivational factors across sexes using the Children's Motivation Analysis Test. Personality and Individual Differences, 10, 483-487.

Boyle, G. J., Start, K. B., & Hall, E. J. (1989). Dimensions of adolescent motivation as measured by higher-order factors in the School Motivation Analysis Test. Journal of School Psychology, 27, 27-33.

Boyle, G. J., Borg, M. G., Falzon, J. M., & Baglioni, A. J. (1993). A structural model of the dimensions of teacher stress: Exploratory and LISREL confirmatory factor analyses. Journal of Educational Psychology (Submitted).

Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern. Psychological Bulletin, 107, 260-273.

Brennan, J. M., & Nitz, L. H. (1986). Stat 1: A statistical toolbox. Kailua, HI: Sugar Mill Software Corp.

Brennan. R. L. (1983). Elements of generalizability theory. Iowa City, IA: American College Testing.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. Journal of Personality, 54, 106-148.

Briggs-Myers, I., & Briggs, K. C. (1985). Myers-Briggs Type Indicator (MBTI). Palo Alto, CA:
Consulting Psychologists Press.

Brody, N. (1988). Personality: In search of individuality. San Diego: Academic

Brody, N. (1992). Intelligence (2nd ed.). New York: Academic.

Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices. British Journal of Mathematical and Statistical Psychology, 37, 1-21.

Browne, M. W. (1990). MUTMUM PC User's guide. Unpublished manuscript. Department of Statistics, University of South Africa, Pretoria.

Buss, A. H. (1989). Personality as traits. American Psychologist, 44, 1378-1388.

Butcher, J. N., Keller, L. S., & Bacon, S. F. (1985). Current developments and future directions in computerized personality assessment. Journal of Consulting and Clinical Psychology, 53, 803-815.

Bynner, J. M., & Romney, D. M. (1986). Intelligence, fact or artefact: Alternative structures for cognitive abilities. British Journal of Educational Psychology, 56, 12-23.

Byrne, B. M. (1988). Testing the factorial validity and invariance of a measuring instrument using LISREL confirmatory factor analyses: A reexamination and application. Multiple Linear Regression Viewpoints, 16, 33-80.

Byrne, B. M. (1989). A primer of LISREL: Basic applications and programming for confirmatory factor analytic models. New York: Springer-Verlag.

Byrne, B. M., & Goffin, R. D. (in press). Modeling MTMM data from additive and multiplicative covariance structures: An audit of construct validity concordance. Multivariate Behavioral Research.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychological Bulletin, 105, 456-466.

Cantor, N., & Kihlstrom, J. F. (1987). Personality and social intelligence.

Englewood Cliffs, NJ: Prentice-Hall.

Carroll, J. B. (1984). Raymond B. Cattell's contributions to the theory of cognitive abilities. Multivariate Behavioral Research, 19, 300-306.

Carroll, J. B. (1991). No demonstration that 'g' is not unitary, but there is more to the story: Comment on Kranzler and Jensen. Intelligence, 15, 423-436.

Cattell, R. B. (1973). Personality and mood by questionnaire. San Francisco: Jossey-Bass.

Cattell, R. B. (1978). The scientific use of factor analysis in behavioral and life sciences. New York: Plenum.

Cattell, R. B. (1979). Personality and learning theory, Vol. 1: The structure of personality in its environment. New York: Springer.

Cattell, R. B. (1980). Personality and learning theory, Vol. 2: A systems theory of maturation and learning. New York: Springer.

Cattell, R. B. (1982a). The development of attribution theory into spectrad theory, using the general perceptual model. Multivariate Behavioral Research, 17, 169-192.

Cattell, R. B. (1982b). The inheritance of personality and ability: Research methods and findings. New York: Academic.

Cattell, R. B. (1983). Structured personality-learning theory: A wholistic multivariate research approach. New York: Praeger.

Cattell, R. B. (1985). Human motivation and the dynamic calculus. New York: Praeger.

Cattell, R. B. (1987a). Intelligence: Its structure, growth, and action. Amsterdam: North-Holland.

Cattell, R. B. (1987b). Psychotherapy by structured learning theory. New York: Springer.

Cattell, R. B. (1992a). Human motivation objectively, experimentally analysed. British Journal of Medical Psychology, 65, 237-243.

Cattell, R. B. (1992b). Superseding the motivational distortion scale. Psychological Reports, 70, 499-502.

Cattell, R. B. (1982c). The psychometry of objective motivation measurement: A response to the critique of Cooper and Kline. British Journal of Educational Psychology, 52, 234-241.

Cattell, R. B., & Johnson, R. C. (1986). (Eds.), Functional psychological testing: Principles and instruments. New York: Brunner/Mazel.

Cattell, R. B., & Krug, S. E. (1986). The number of factors in the 16PF: A review of the evidence with special emphasis on methodological problems. Educational and Psychological Measurement, 46, 509-522.

Cattell, R. B., & Schuerger, J. M. (1978). Personality theory in action: Handbook for the O-A (Objective-Analytic) Test Kit. Champaign, IL: Institute for Personality and Ability Testing.

Ceci, S. J. (1990). On intelligence...More or less. Englewood Cliffs, NJ: Prentice Hall.

Child, D. (1990). The essentials of factor analysis (2nd ed.). London: Cassell.

Clark, L. A., & Livesley, W. J. (in press). Two approaches to identifying the dimensions of personality disorder: Convergence on the five-factor model. In P. T. Costa, Jr., & T. A. Widiger (Eds.), Personality disorders and the five-factor model of personality. Washington, DC: American Psychological Association.

Clark, L. A., McEwen, J. L., Collard, L. M., & Hickok, L. G. (1993). Symptoms and traits of personality disorder: Two new methods for their assessment. Psychological Assessment, 5, 81-91.

Clark, L. A., Vorhies, V., & McEwen, J. L. (in press). Personality disorder symptomatology from the five-factor perspective. In P. T. Costa, Jr., & T. A. Widiger (Eds.), Personality disorders and the five-factor model of personality. Washington, DC: American Psychological Association.

Cole, D. A., & Maxwell, S. E. (1985). Multitrait-multimethod comparisons across populations: A confirmatory factor analytic approach. Multivariate Behavioral Research, 20, 389-417.

Comrey, A. L. (1980). Handbook of interpretation for the Comrey Personality Scale. San Diego: Educational & Industrial Testing Service.

Comrey, A. L., & Lee, H. B. (1992). A first course in factor analysis (rev. 2nd ed.). Hillsdale, NJ:Erlbaum.

Conley, J. J. (1984). Longitudinal consistency of adult personality: Self-reported psychological characteristics across 45 years. Journal of Personality and Social Psychology, 47,
1325-1334.

Connell, J. P. (1987). Structural equation modeling and the study of child development: A question of goodness of fit. Child Development, 58, 167-175.

Costa, P. T., & McCrae, R. R. (1992a). Four ways five factors are basic. Personality and Individual Differences, 13, 653-665.

Costa, P. T., & McCrae, R. R. (1992b). Normal personality assessment in clinical practice: The NEO Personality Inventory. Psychological Assessment, 4, 5-13.

Costa, P. T., & McCrae, R. R. (1992c). The NEO Personality Inventory (Revised) manual. Odessa, FL: Psychological Assessment Resources.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart & Winston.

Cronbach, L. J. (1990). Essentials of psychological testing (5th ed.). New York: Harper & Row.

Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures analysis and the "problem" of sample size: A clarification. Psychological Bulletin, 109, 512-519.

Cuttance, P. & Ecob, R. (1987). (Eds.), Structural modeling by example: Applications in educational, sociological, and behavioral research. New York: Cambridge.

Digman, R. M. (1990). Personality structure: Emergence of the five-factor model. Annual Reviewof Psychology, 41, 417-440.

Eaves, L. J., Eysenck, H. J., & Martin, N. G. (1989). Genes, culture and personality. London: Academic.

Elliott, C. D., Murray, D. J., & Pearson, L. S. (1983). British Ability Scales (2nd ed.). London: Nelson.

Eysenck, H. J. (1984). Cattell and the theory of personality. Multivariate Behavioral Research, 19,
323-336.

Eysenck, H. J. (1985a). Decline and fall of the Freudian empire. London: Viking.

Eysenck, H. J. (1985b). Review of the California Psychological Inventory. In J. V. Mitchell (Ed.), Ninth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements.

Eysenck, H. J. (1990a). Biological dimensions of personality. In L. A. Pervin (Ed.), Handbook of personality: Theory and research. New York: Guilford.

Eysenck, H. J. (1990b). Genetic and environment contributions to individual differences: The three major dimensions of personality. Journal of Personality, 98, 245-261.

Eysenck, H. J. (1991). Dimensions of personality: 16, 5, 3? Personality and Individual Differences, 12, 773-790.

Eysenck, H. J. (1992). Four ways five factors are not basic. Personality and Individual Differences, 13, 667-673.

Eysenck, H. J., & Eysenck, M. W. (1985). Personality and individual differences. New York: Plenum.

Eysenck, M. W., & Eysenck, H. J. (1980). Mischel and the concept of personality. British Journal of Psychology, 71, 191-204.

Fernandez, E. (1990). Artifact in pain ratings, its implications for test-retest reliability, and correction by a new scaling procedure. Journal of Psychopathology and Behavioral Assessment, 12, 1-15.

Fernandez, E., Nygren, T. E., & Thorn, B. E. (1991). An open-transformed scale for correcting ceiling effects and enhancing retest reliability: The example of pain. Perception and Psychophysics, 49, 572-578.

Friedman, A., Webb, J., & Lewark, R. (1989). Psychological testing with the MMPI. London: Erlbaum.

Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of typical engagement. Journal of Educational Psychology, 84, 537-552.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. Psychological Assessment, 4, 26-42.

Goldstein, H. (1980). Dimensionality, bias, independence and measurement scale problems in latent trait test score models. British Journal of Mathematical and Statistical Psychology, 33, 234-246.

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Erlbaum.

Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. Multivariate Behavioral Research, 25, 33-39.

Gorsuch, R. L. (1991). UniMult guide (Version 1.1). Altadena, CA: UniMult.

Gough, H. G. (1987). California Psychological Inventory (CPI) (rev. ed.). Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G. (1989). Review of the Objective-Analytic (O-A) Test Battery. The tenth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements.

Grayson, D. A. (1986). Latent trait analysis of the Eysenck Personality Questionnaire. Journal of Psychiatric Research, 20, 217-235.

Gruijter, de D. N. M., & Kamp, van der L. J. T. (1990). Generalizability theory. In R. K. Hambleton & J. N. Zaal (Eds.), Advances in educational and psychological testing: Theory and applications. Boston: Kluwer.

Guilford, J. P. (1975). Factors and factors of personality. Psychological Bulletin, 82, 802-814.

Guilford, J. P. (1981). Higher-order structure-of-intellect abilities. Multivariate Behavioral Research, 16, 411-435.

Guilford, J. P. (1985). The structure-of-intellect model. In B. B. Wolman (Ed.), Handbook of intelligence: Theories, measurements and applications. New York: Wiley.

Guthrie, G. (1985). Review of the Clinical Analysis Questionnaire. In J. V. Mitchell, Jr. (Ed.), Ninth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements.

Guttman, L. (1992). The irrelevance of factor analysis for the study of group differences. Multivariate Behavioral Research, 27, 175-204.

Hakstian, A. R., & Cattell, R. B. (1982). The Comprehensive Ability Battery. Champaign, IL: Institute for Personality & Ability Testing.

Hakstian, A. R., & Woolsey, L. K. (1985). Validity studies using the comprehensive ability battery (CAB): IV. Predicting achievement at the university level. Educational and Psychological Measurement, 45, 329-341.

Hakstian, A. R., Rogers, W. T., & Cattell, R. B. (1982). The behavior of number-of-factors rules with simulated data. Multivariate Behavioral Research, 17, 193-219.

Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston: Kluwer.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Newbury Park, CA: Sage.

Hartmann, W. M. (1990). ProcCALIS: Analysis of covariance structures. In F. Faulbaum, R. Haux & K. Jöckel (Eds.), SoftStat '89. Fortschritte der Statistik--Software 2 (pp. 74-81). Stuttgart: Gustav Fischer Verlag.

Hayduk, L. A. (1987). Structural equation modeling with LISREL: Essentials and advances. Baltimore: Johns Hopkins University Press.

Heck, van G. L. M. (1988). Modes and models in anxiety. Anxiety Research, 1, 199-214.

Helmes, E., & Reddon, J. R. (1993). A perspective on developments in assessing psychopathology: A critical review of the MMPI and MMPI-2. Psychological Bulletin, 113, 453-471.

Hessler, G. L. (1982). Use and interpretation of the Woodcock-Johnson Psycho-Educational Battery. Allen, TX: DLM Teaching Resources.

Hogan, R. (1986). Manual for the Hogan Personality Inventory (HPI). Minneapolis, MN: National Computer Systems.

Holden, R. R., Reddon, J. R., Jackson, D. N., & Helmes, E. (1983). The construct heuristic applied to the measurement of psychopathology. Multivariate Behavioral Research, 18, 37-46.

Holden, R. R., Kroner, D. G., Fekken, G. C., & Popham, S. M. (1992). A model of personality test item response dissimulation. Journal of Personality and Social Psychology, 63, 272-279.

Holland, J. L. (1985). Making vocational choices: A theory of vocational personalities and work environments. Englewood-Cliffs, NJ: Prentice-Hall.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), Handbook of multivariate experimental psychology. New York: Plenum.

Horn, J. L., & Stankov, L. (1980). Auditory and visual factors of intelligence. Intelligence, 6, 165-185.

Horn, J. L., & Stankov, L. (1982). Comments about a chameleon theory: Level I/Level II. Journalof Educational Psychology, 74, 874-878.

Howard, R. W. (1993). On what intelligence is. British Journal of Psychology, 84, 27-37.

Hunter, J. E., Schmidt, F. L., & Rauschenberger, J. M. (1984). Methodological, statistical, and ethical issues in the study of bias in psychological tests. In C. E.

Reynolds & R. T. Brown (Eds.), Perspectives on bias in mental testing (pp. 41-99). New York: Plenum.

Hutchinson, T. P. (1991). Controversies in item response theory. Sydney: Rumsby Scientific.

Jensen, A. R. (1991). General mental ability: From psychometrics to biology. Diagnostique, 16, 134-144.

Jinks, J. L., & Fulker, J. W. (1970). Comparison of the biometrical, genetical, MAVA and classical approaches to the study of human behaviour. Psychological Bulletin, 73, 311-349.

John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. Pervin (Ed.), Handbook of personality theory and research. (pp. 66-100). New York: Guilford.

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. European Journal of Personality, 2, 171-203.

John, O. P., Hampson, S. E., & Goldberg, L. R. (1991). The basic level in personality-trait hierarchies: Studies of trait use and accessibility in different contexts. Journal of Personality and Social Psychology, 60, 348-361.

Johnson, R. C. (1986). Personality assessment by observers in normal and psychiatric data. In R. B. Cattell & R. C. Johnson (Eds.), Functional psychological testing: Principles and instruments. New York: Brunner/Mazel.

Jöreskog, K. G., & Sörbom, D. (1988). PRELIS: A program for multivariate data screening and data summarization: A preprocessor for LISREL. Mooresville, IN: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (1989). LISREL 7: A guide to the program and applications (2nd ed.). Chicago: SPSS Inc.

Kameoka, V. A. (1986). The structure of the Clinical Analysis Questionnaire and depression symptomatology. Multivariate Behavioral Research, 21, 105-122.

Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. Multivariate Behavioral Research, 25, 137-155.

Kaufman, A. S. (1990). Assessing adolescent and adult intelligence. Chicago: Allyn

& Bacon.

Kaufman, A. S., & Kaufman, N. L. (1983). Kaufman Assessment Battery for Children (K-ABC). Circle Pines, MN: American Guidance Service.

Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: CBS.

Kline, P. (1980). The psychometric model. In A. J. Chapman & D. M. Jones (Eds.)., Models of man. Leicester: The British Psychological Society.

Kline, P. (1986). A handbook of test construction: Introduction to psychometric design. London: Methuen.

Kline, P. (1993a). Personality: The psychometric view. London: Routledge.

Kline, P. (1993b). The handbook of psychological testing. London: Routledge.

Kline, P., & Cooper, C. (1984). A construct validation of the Objective-Analytic Test Battery (OATB). Personality and Individual Differences, 5, 323-337.

Kline, P., & Cooper, C. (1984). The factor structure of the Comprehensive Ability Battery. British Journal of Educational Psychology, 54, 106-110.

Kranzler, J. H., & Jensen, A. R. (1991). The nature of psychometric 'g': Unitary process of a number of independent processes? Intelligence, 15, 379-422. Personality and Ability testing.

Krug, S. E. (1980). Clinical Analysis Questionnaire manual. Champaign, IL: Institute for Personality & Ability Testing.

Krug, S. E., & Johns, E. F. (1986). A large scale cross-validation of second-order personality structure defined by the 16PF. Psychological Reports, 59, 683-693.

Livneh, H., & Livneh, C. (1989).The five-factor model of personality: Is evidence of its cross-measure validity premature? Personality and Individual Differences, 10, 75-80.

Loehlin, J. C. (1990). Component analysis versus common factor analysis: A case of disputed authorship. Multivariate Behavioral Research, 25, 29-31.

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. Applied Psychological Measurement, 13, 335-361.

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. Applied Psychological

Measurement, 15, 47-70.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103, 391-410.

Martin, J. A. (1987). Structural equation modeling: A guide for the perplexed. Child Development, 58, 33-37.

Masson, J. M. (1990). Final analysis: The making and unmaking of a psychoanalyst. London: Fontana.

Mathews, G. (1989). The factor structure of the 16PF: Twelve primary and three secondary factors. Personality and Individual Differences, 10, 931-940.

McArdle, J. J. (1984). On the madness in his method: R. B. Cattell's contributions to structural equation modeling. Multivariate Behavioral Research, 19, 245-267.

McArdle, J. J. (1990). Principal versus principals of structural factor analyses. Multivariate Behavioral Research, 25, 81-87.

MacCallum, R. (1985). Some problems in the process of model modification in covariance structure modeling. Paper presented to the European Meeting of the Psychometric Society, Cambridge, England (July).

MacCallum, R. (1986). Specification searches in covariance structure modeling. Psychological Bulletin, 100, 107-120

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. Journal of Personality and Social Psychology, 52, 81-90.

McCrae, R. R., & Costa, P. T. (1989). The structure of interpersonal traits: Wiggins circumplex and the five-factor model. Journal of Personality and Social Psychology, 56, 586-595.

McDonald, R. P. (1985). Factor analysis and related methods. Hillsdale, NJ: Erlbaum.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness-of-fit. Psychological Bulletin, 107, 247-255.

McKenzie, J. (1988). Three superfactors in the 16PF and their relation to Eysenck's P, E and N. Personality and Individual Differences, 9, 843-850.

Mershon, B., & Gorsuch, R. L. (1988). Number of factors in the personality sphere: Does increase in factors increase predictability of real-life criteria. Journal of Personality and Social Psychology, 55, 675-680.

Messick, S. (1992). Multiple intelligence or multilevel intelligence? Selective emphasis on distinctive properties of hierarchy: On Garner's *Frames of mind* and Sternberg's *Beyond IQ* in the context of theory and research on the structure of human abilities. Psychological Inquiry, 3, 365-384.

Miller, K. M. (1988). The analysis of personality in research and assessment: In tribute to Raymond B. Cattell. London: Independent Assessment and Research Centre.

Millsap, R. E. (1990). A cautionary note on the detection of method variance in multitrait-multimethod data. Journal of Applied Psychology, 75, 350-353.

Mischel, W. (1984). Convergencies and challenges in the search for consistency. American Psychologist, 39, 351-364.

Moos, R. (1985). Perimenstrual symptoms: A manual and overview of research with the Menstrual Distress Questionnaire. Palo Alto, CA: Department of Psychiatry and Behavioral Sciences, Stanford University.

Morey, L. C. (1991). Personality Assessment Inventory: Professional manual. Odessa, FL: Psychological Assessment Resources.

Mulaik, S. A. (1986). Factor analysis and psychometrika: Major developments. Psychometrika, 51, 23-33.

Mulaik, S. A. (1987). Toward a conception of causality applicable to experimentation and causal modeling. Child Development, 58, 18-32.

Mulaik, S. A. (1990). Blurring the distinctions between component analysis and common factor analysis. Multivariate Behavioral Research, 25, 53-59.

Mulaik, S. A., James, L. R., Alstine, van J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, 105, 430-445.

Muthén, B. O. (1988). LISCOMP: Analysis of linear structural equations with comprehensive measurement model (2nd ed.). Mooresville, IN: Scientific Software.

Nesselroade, J. R., & Cattell, R. B. (1988). (Eds.). Handbook of multivariate experimental psychology (rev. 2nd ed.). New York: Plenum.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes:

Replicated factor structure in peer nomination personality ratings. Journal of Abnormal and Social Psychology, 66, 574-583.

O'Toole, B. I., & Stankov, L. (1992). Ultimate validity of psychological tests. Personality and Individual Differences, 13, 699-716.

Poon, W. Y., & Lee, S. Y. (1987). Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. Psychometrika, 52, 409-430.

Powell, G. E. (1979). Brain and personality. Farnborough, UK: Saxon House.

Prosser, R., Rasbash, J., & Goldstein, H. (1991). ML3 software for three-level analysis: User's guide for V.2. London: Institute of Education, University of London.

Romney, D. M., & Bynner, J. M. (1992). The structure of personal characteristics. Westport, CT: Praeger.

Roseboom, W. W. (1990). Whatever happened to broad perspective? Multivariate Behavioral Research, 25, 61-65.

Rudy, T. E., Turk, D. C., & Brody, M. C. (1992). Quantification of biomedical findings in chronic pain: Problems and solutions. In D. Turk & R. Melzack (Eds.), Handbook of pain assessment. New York: Guilford.

Russell, E. W., Neuringer, C., & Goldstein, G. (1970). Assessment of brain damage: A neuropsychological key approach. New York: Wiley-Interscience.

Schmidt, L. R. (1988). Objective personality tests: Some clinical applications. In K. M. Miller (Ed.), The analysis of personality in research and assessment: In tribute to Raymond B. Cattell. London: Independent Assessment & Research Centre.

Schmitt, M. J., & Steyer, R. (1993). A latent state-trait model (not only) for social desirability. Personality and Individual Differences, 14, 519-529.

Schmitt, N., & Stults, D. N. (1986). Methodology review: Analysis of multitrait-multimethod matrices. Applied Psychological Measurement, 10, 1-22.

Schönemann, P. H. (1990). Facts, fictions, and common sense about factors and components. Multivariate Behavioral Research, 25, 47-51.

Schuerger, J. M. (1986). Personality assessment by objective tests. In R. B. Cattell & R. C. Johnson (Eds.), Functional psychological testing: Principles and instruments. New York: Brunner/Mazel.

Schuerger, J. M. (1992). The Sixteen Personality Factor Questionnaire and its junior versions. Journal of Counseling and Development, 71, 231-244.

Schwartz, B., & Reisberg, D. (1992). Learning and memory. New York: Norton.

Shavelson, R., Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. American Psychologist, 44, 922-932.

Snook, S. C., & Gorsuch, R. L. (1989). Principal component analysis versus common factor analysis: A Monte Carlo study. Psychological Bulletin, 106, 148-154.

Snow, R. E., Killonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence (Vol. 2). Hillsdale, NJ: Erlbaum.

Soldz, S., Budman, S., Demby, A., & Merry, J. (1993). Representation of personality disorders in circumplex and five-factor space: Explorations with a clinical sample. Psychological Assessment, 5, 41-52.

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1986). Manual for the State-Trait Anxiety Inventory-Form Y. Palo Alto: Consulting Psychologists Press.

Stankov, L. (1978). Fluid and crystallized intelligence and broad perceptual factors among 11 to 12 year olds. Journal of Educational Psychology, 70, 324-334.

Stankov, L. (1980). Psychometric factors as cognitive tasks: A note on Carroll's "New Structure of Intellect." Intelligence, 4, 65-71.

Stankov, L. (1983). Attention and intelligence. Journal of Educational Psychology, 75, 471-490.

Stankov, L. (1986). Kvashchev's experiment: Can we boost intelligence? Intelligence, 10, 209-230.

Stankov, L. (1987). Level I/II: A theory ready to be archived. In S. Modgil & C. Modgil (Eds.), Arthur Jensen: Consensus and controversy (Ch. 4). London: Falmers Press.

Stankov, L. (1989). Attentional resources and intelligence: A disappearing link. Personality and Individual Differences, 10, 957-968.

Stankov, L., & Chen, k. (1988). Can we boost fluid and crystallized intelligence? A structural modeling approach. Australian Journal of Psychology, 40, 363-376.

Stankov, L., & Crawford, J. D. (1983). Ingredients of complexity in fluid

intelligence. Learningand Individual Differences, 5, 73-111.
Stankov, L., & Cregan, A. (1993). Quantitative and qualitative properties of an intelligence test: Series Completion. Learning and Individual Differences, in press.
Stankov, L., & Horn, J. L. (1980). Human abilities revealed through auditory tests. Journal of Educational Psychology, 72, 19-42.
Stankov, L., & Myors, B. (1990). The relationship between working memory and intelligence: Regression and COSAN analyses. Personality and Individual Differences, 11, 1059-1068.
Stankov, L., Horn, J. L., & Roy, T. (1980). On the relationship between Gf/Gc theory and Jensen's Level I/Level II theory. Journal of Educational Psychology, 72, 796-809.
Steiger, J. H. (1989). EzPATH: A supplementary module for SYSTAT and SYGRAPH. Evanston, IL: SYSTAT Inc.
Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. Child Development, 58, 134-146.
Thorndike, R. L. (1982). Applied psychometrics. Boston: Houghton Mifflin.
Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). Technical manual: Stanford-Binet Intelligence Scale (4th ed.). Chicago: Riverside.
Tupes, E. C., & Christal, R. E. (1961). Recurrent personality factors based on trait ratings. USAF ASD Technical Report, No. 61-97.
Velicer, W. F., & Jackson, D. N. (1990a). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. Multivariate Behavioral Research, 25,
1-28.
Velicer, W. F., & Jackson, D. N. (1990b). Component analysis versus common factor analysis: Some further observations. Multivariate Behavioral Research, 25, 97-114.
Wainer, H., Dorans, N. J., Flauger, R., Green, B. F., Mislevy, R. J., Steinberg, L, & Thissen, D. (1990). Computerised adaptive testing: A primer. Hillsdale, NJ: Erlbaum.
Webb, N. M., Shavelson, R. J., & Maddahian, E. (1983). Multivariate generalizability theory.
In L. J. Fyans, Jr. (Ed.), New directions for testing and measurement:

Generalizability theory (pp. 67-82). San Francisco: Jossey-Bass.

Wechsler, D. (1991). Wechsler Intelligence Scale for Children-III. San Antonio, TX: Psychological Corporation.

Weiss, D. J., & Yoes, M. E. (1990). Item response theory. In R. K. Hambleton & J. N. Zaal (Eds.), Advances in educational and psychological testing: Theory and applications. Boston: Kluwer.

Widaman, K. F. (1990). Bias in pattern loadings represented by common factor analysis and component analysis. Multivariate Behavioral Research, 25, 89-95.

Wiggins, J. S., & Trapnell, P. D. (in press). Personality structure: The return of the Big Five. In
S. R. Briggs, R. Hogan, & W. H. Jones (Eds.), Handbook of personality psychology. Orlando, FL: Academic.

Wiggins, J. S., & Pincus, A. L. (1992). Personality: Structure and assessment. Annual Review of Psychology, 43, 473-504.

Williams, L. J., Cote, J. A., & Buckley, M. R. (1989). Lack of method variance in self-reported affect and perceptions at work: Reality or artifact? Journal of Applied Psychology, 74,
462-468.

Winer, B. J., Brown, D. R., Michels, K. M. (1991). Statistical principles in experimental design. (rev. 3rd ed.). New York: McGraw-Hill.

Woodcock, R. W., & Mather, N. (1989). Woodcock-Johnson Tests of Cognitive Ability. Alen, TX: DLM Teaching Resources.

Woliver, R. E., & Saeks, S. D. (1986). Intelligence and primary aptitudes: Test design and tests available. In R. B. Cattell & R. C. Johnson (Eds.), Functional psychological testing: Principles and instruments. New York: Brunner/Mazel.

Zuckerman M. (1991). Psychobiology of personality. New York: Cambridge University Press.

Zuckerman, M. (1992). What is a basic factor and which factors are basic? Turtles all the way down. Personality and Individual Differences, 6, 675-681.

Zuckerman, M., Kuhlman, D. M., & Camac, C. (1988). What lies beyond E and N? Factor analyses of scales believed to measure basic dimensions of personality. Journal of Personality and Social Psychology, 54, 96-107.

Zuckerman, M., Kuhlman, D. M., Thornquist, M., & Kiers, H. (1991). Five (or three) robust questionnaire scale factors of personality without culture. Personality and Individual Differences, 12, 929-941.

ENDNOTES

1. Factor analysis is a mathematico-statistical procedure which is applied to an intercorrelation matrix, with the goal of delineating the underlying (often "causal") dimensions (latent traits or factors), responsible for the observed correlations between a larger number of variables.

2. Cattell regards the emphasis on higher-stratum factors as problematic. According to him, the ancient Greeks started with four elements--air, earth, fire and water--but modern chemists recognize the need for 100 elements. Popularity of three or five factors, like the above analogy, is an understandable but inadequate view of the world. This has been shown by the recent wide survey of predictions of occupational and clinical performances in which prediction from 16 factors greatly exceeded that from three or five factors (Mershon & Gorsuch, 1988). One could deduce this also from the Cattell-White formula $V_2 = V_1 V_{12}$ --in which the loadings of secondaries on items (and life performances) are products of fractions. Secondaries are relatively useless tools, except to give situational effects to the primaries (Birkett-Cattell, 1989).

3. Since Cattellian psychometric instruments are constructed factor analytically, the primary factors (abilities, traits, dynamics, and states) represent underlying "causal" and psychologically meaningful dimensions, such that the validity of the Cattellian psychometric model (including both the ability and personality submodels) can be quantitatively measured and therefore tested empirically.