Humanities & Social Sciences papers   Faculty of Humanities and Social Sciences

2009

Methods of personality assessment

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

Edward Helmes

Follow this and additional works at: http://epublications.bond.edu.au/hss_pubs

Part of the Personality and Social Contexts Commons

Word Count = 8,206

Methods of Personality Assessment

Gregory J. Boyle

Department of Psychology, Bond University
and
Department of Psychiatry, University of Queensland

&

Edward Helmes

Department of Psychology, James Cook University

This chapter cannot provide an exhaustive review of the many approaches to personality

assessment that are in common use today because of the vast size of the area. With

entire books devoted to individual instruments, a brief chapter such as this is necessarily

limited in its scope. In particular, the chapter will not address methods of projective

personality assessment. Those interested in an introduction to such methods may

consult the relevant chapters in books by Groth-Marnat (2003), and Weiner (1997), as

well as the commentaries in the *Journal of Personality Assessment* relating to the use of

projective instruments such as the *Rorschach Inkblot Test*. For a critical perspective,

readers may consult, for example, Hunsley, Lee, and Wood (2003).

In the present chapter, the approach being taken focuses on contrasting the

multidimensional personality assessment instruments constructed using factor analysis

by Raymond B. Cattell and his colleagues with those multidimensional scales developed using other approaches for assessing personality attributes, notably the construct-oriented methods advocated by Douglas Jackson. While other measurement approaches are available, subjective self-report instruments remain the dominant form of personality assessment, whether administered on a computer screen and scored online (e.g., see Drasgow & Olson-Buchanan, 1999), or in a more traditional answer sheet and question booklet, or combined question and answer sheet format. Their economy, apparent ease of use and interpretation, and freedom of the need for trained interviewers (or even third parties in some cases) provide advantages that often outweigh benefits of other approaches to personality assessment. Although other methods, including objective tests (T-data) measures such as the *Objective Analytic Battery* (Cattell & Schuerger, 1978; Schuerger, 1986) are mentioned briefly, the bulk of this chapter will be concerned with the popular self-report techniques.

For many users of personality assessment instruments, test construction methods are irrelevant or of little interest. The methods used do, however, directly relate to the content of measures and influence how the instruments should be interpreted. The assumptions made and the choice of procedures used during the process of refining and selecting items very much determine the final product. For example, concerns for reading level can influence item phrasing and the complexity of ideas being expressed, while assumptions about sex differences will influence whether scales have gender-based norms or not, thus determining the nature of inferences that can be drawn about clients. The work of Cattell and his colleagues is notable for its empirical use of the

inductive-hypothetico-deductive factor-analytic method used to identify items that were deemed to reflect personality traits as expressed and structured in the trait lexicon of the English language. Cattell's pioneering psychometric research into the structure and measurement of human personality also involved several media of measurement (L-data, Q-data, and T-data), and did not restrict itself to self-report methodology alone (see Boyle, 2006). In fact, Cattell is listed (along with Freud, Piaget, and Eysenck) among the 10 most highly cited psychologists of the 20th century as indexed in the published journal literature (Haggbloom et al., 2002).

In contrast, Jackson's (2000) work in the area of personality assessment emerged from a rather different orientation, starting in clinical psychology and being influenced by Gardner Murphy and by the seminal papers of Cronbach and Meehl (1955), Campbell and Fiske (1959), and Loevinger (1957). Jackson's interest in multivariate assessment and factor-analytic methods rivaled Cattell's, but his approach to test construction relied upon there being substantive theory in place that guided the processes of item writing and item analysis. He also valued the multitrait-multimethod approach to construct validity and this technique formed part of the development process for many of his instruments. While both these psychometricians utilized exploratory factor analysis, their emphasis on it differed. Cattell argued that when properly used, the method provided insights into the natural structure of personality and he used it as the primary method to form scales from sets of items. Cattell (e.g., 1973; Burdsal & Vaughn, 1974) argued against factor analyzing items and for the use of homogeneous groups of items (*item parcels*). Later Comrey (1988) also argued for the use of sets of items, which he

termed *Factored Homogeneous Item Dimension*s (FHIDs; Comrey, 1967, 1984), and not

individual items as the input to factor analysis. In contrast, Jackson used factor analysis

more as a method of confirming structures that had been developed on theoretical

grounds, but he often relied more upon basic correlational analyses of item pools than

upon item factor analyses.

Briggs and Cheek (1986) provided an overview of the relevant issues that need to be

considered in using factor analysis for scale construction and give examples of where

factor analysis can help refine a scale intended to measure a single construct and where it

can obscure matters. They noted the increasingly important distinction between

e*xploratory* methods, in which the goal in item factor analysis is to identify any structure

underlying a set of items, and *confirmatory* methods in which the goal is to verify if a

theoretical or predetermined structure is indeed supported in the set of items. Item factor

analysis involves several complex issues (including the lower reliability of individual

items) and there are now several alternative computational methods for analyzing

personality questionnaire items (see Wirth & Edwards, 2007 for a review of current

methods).

Exploratory factor-analytic (EFA) methodology has progressed since the publication

of Cattell's (1978) treatise (for a detailed discussion of EFA methodology, see Boyle,

Stankov, & Cattell, 1995; Gorsuch, 1983, 1988; McDonald, 1985). With the advent

of more powerful and cheaper computers, newer techniques including confirmatory

factor analysis or CFA (Mulaik, 1988), as one application of structural equation

modeling (SEM) implemented via LISREL, EQS or other similar computer

programs, have become commonplace (Bentler, 1988). Over the past three decades,

issues related to EFA have continued to simmer without a clear consensus upon

several major issues that involve the commonly used methods. Historically, Cattell

(1978, 1988) favoured the traditional common factor-analytic model for constructing

and evaluating personality instruments. In contrast, the crude principal components

analysis method, based on a less sophisticated underlying mathematical model,

because of its ease of application, became in practice, the most commonly used

procedure. The various issues on this topic were reviewed by Velicer and Jackson

(1990) and an extensive series of comments and rebuttals followed in that issue of

Multivariate Behavioral Research that suggested that no clear consensus among

experts was evident on several basic matters.

One unfortunate aspect of the differing perspectives on the best version of factor

analysis is that various EFA methods are frequently used without a full understanding

of their limitations, as are the exploratory applications of SEM procedures. When

properly applied, the various CFA procedures that are in common use can be quite

powerful and informative, but they also have their shortcomings, particularly when

they are used for exploratory analyses. Cuttance (1987, p. 243) commented on such

applications with one example of the pitfalls in using structural equation models for

exploratory ends:

"MacCallum (1985) investigated the process of the exploratory

fitting of models in simulated data…for which the true model was

known. He found that only about half of the exploratory searches

located the true model.…He obtained this limited rate of

success…in samples of 300 observations…and his success rate in

smaller samples (N=100) was zero….An exploratory analysis of

data thus entails the risk of inducing an interpretation founded on the

idiosyncracies of individual samples."

Wirth and Edwards (2007) also noted that most SEM programs require substantial

sample sizes and cannot handle the number of parameters required if large multiscale

personality measures are analyzed at the item level. Most item response theory (IRT)

programs can deal with the number of items, but may encounter difficulties if the

assumption of homogeneous scales is violated, which is likely to happen with many

personality constructs.

In order to enable comprehensive assessment and to undertake multidimensional

measurement of personality traits, it is generally considered desirable to use a variety of

measurement media, including subjective questionnaires, structured interviews and

objective test instruments (Cattell, 1986a, b; Smith, 1988). Cattell argued cogently that

personality traits should be identified through multiple measurement media including

subjective ratings or life-record data (L-Data), subjective self-report questionnaire data

(Q-data), and objective test data (T-data). The choice of specific media of measurement has critically important implications in terms of susceptibility to response distortion, as well as for psychometric properties such as reliability and validity (Cattell, 1986a, c, d). These issues of response distortion were also addressed in more depth by other workers, including those who addressed the specific question of response styles such as social desirability and acquiescence in the MMPI and other self-report instruments, beginning with Alan Edwards (1957) and finalized (at least in the eyes of those who preferred to minimize the stylistic responding argument) by Jack Block (1965). The importance of social desirability as an alternative explanation for certain results or as a confounding variable has faded somewhat in visibility in personality assessment since the major debates on the topic of the 1960s and 70s. At the same time, there is widespread recognition of the nature of social desirability as a measurement confound and as an important personality variable in itself (Helmes, 2000). The latter position is illustrated, for example, by the work of Paulhus (1984, 1986) and the development of his measures of self-deception and impression management as the major dimensions of social desirability (Paulhus, 1998).

The types of information referred to by Cattell as L-data are more commonly known as *biodata*, and includes such biographical information as education, work experience, and volunteer activities. Biodata have been used in various applications (Stokes, Mumford, & Owens, 1994), but are most widely used in personnel selection in industry and the military. Some research (Mount, Witt, & Barrick, 2000) suggests that biographical data appear to account for additional variance over and above that measured by self-report personality measures and general mental ability, supporting Cattell's argument that L-

data can provide valuable information beyond that obtained with conventional personality measures.

Based on the assumption that personality characteristics are represented linguistically, a major Q-data instrument, the *Sixteen Personality Factor Questionnaire* (16PF) was constructed using methodologically-sound factor analytic procedures by Cattell and his colleagues (e.g., Cattell, Eber, & Tatsuoka, 1970; Cattell, 1994; Cattell & Krug, 1986; Krug, 1981). The 16PF was based on exploratory factor analyses of several clusters of personality traits that had been derived from a comprehensive search of over 4000 trait-descriptive terms relating to personality in the English language compiled by Allport and Odbert (1936). This work represented a significant development in both its scope and its reliance upon factor analysis.

Psychological instruments can of course be developed by other methods than the use of factor analysis. Burisch (1984) classed factor-analytic methods as *inductive*, in contrast to external methods such as the criterion groups method used for the original development of the Minnesota Multiphasic Personality Inventory (MMPI, Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). Burisch classed Jackson's construct-oriented approach as *deductive*, along with other methods that can be described as rational, intuitive, or theoretical. Burisch's (1984) review of the relevant literature suggested that there were no demonstrable differences in validity correlations among scales constructed using different methodological approaches. Later research suggested

that some methods do lead to higher validity coefficients than others. For example, a subsequent study using the traits of extraversion and dominance contrasted Jackson's (1984) *Personality Research Form* (PRF) and 16PF 4th edition (16PF4) in predicting job performance of a group of 487 managers and did find some differences (Goffin, Rothstein, & Johnston, 2000). They concluded that the construct-based PRF had "distinct predictive advantages" (p. 261) over the 16PF4 in that context, but also noted that there was very still little relevant literature on such comparisons. Hough and Paullin (1994) also concluded that there were some benefits of the construct-based method of scale development, as did Burisch (1986) in his later comparison of methods of scale construction. What limited evidence exists thus suggests that the method of test construction may well have implications for the validity of personality measures.

The development of multiscale measurement instruments has become a multistage process, and it is noteworthy that many of the major personality instruments in use today were developed decades ago. The actual procedures used go beyond the simple classifications used to compare strategies that were used by Burisch (1984, 1986) and others. The methods used can be illustrated by the procedures outlined by Jackson (1970), who advocated a sequential process that stressed the importance of convergent and divergent validity and the suppression of confounding response styles. Factor analysis could be used at different stages in the process, but it was always secondary to substantive considerations. These procedures were used in his measures of normal personality, the *Personality Research Form* (Jackson, 1984) and *Jackson Personality Inventory* (Jackson, 1994). Interestingly, the latest (fifth) edition of the 16PF (Cattell,

Cattell, & Cattell, 1994) adopted item analytic approaches similar to those used by

Jackson in order to promote convergent properties among the items of given scales.

One important issue in evaluating personality assessment instruments relates to the

actual items and how they are phrased. Emphasis on this aspect of instruments has been

more evident in the area of public opinion and attitude measurement (Sudman &

Bradburn, 1983; Schwarz & Sudman, 1996), but has also been an issue with personality

measures (e.g., Nicholls, Licht, & Pearl, 1982) and with life history (L-data) in the form

of symptoms (Schwarz, 1999). Care on these matters is important for defining both how

a particular scale measures what it is designed to measure (Angleitner, John, & Lohr,

1986) and what it is not intended to measure, which are aspects of convergent and

discriminant validity. Angleitner et al. rated various item characteristics and noted that

ostensibly parallel forms had some scales with different item properties across forms,

and that item complexity frequently affected the ability to immediately understand the

item. They asserted that all but two of the 16PF A and B scales had more than 50% of

items with poor understandability or high ambiguity. This illustrates how sequential test

construction strategies that incorporate structured analysis of item properties in the early

stages can lead to better quality items for final testing. Such strategies can also build in

analyses to foster convergent and discriminant properties in the scales. Rudinger and

Dommel (1986) provided an analysis of a multitrait-multimethod analysis of the German

translation of Jackson's PRF that illustrates both how such analyses can be performed,

but also the properties of the instrument itself.

Within the normal personality trait domain, potentially controversial items pertaining to religion, sex, and politics are often excluded from questionnaires. The presence of such items was one factor leading to the revision of the MMPI (Butcher et al., 1989). Likewise, items relating to social desirability and other response sets typically are kept at a minimum during the developmental stages for large, multiscale measures. This may not be the case for shorter, more narrowly focused measures of single or a few attributes where extensive developmental research on a measure may not have been completed before an instrument appeared in the research literature. Users of such instruments should always investigate the methods by which a scale was developed and the reported psychometric properties of personality measures before using them, but detailed information on the development of scales may not be widely available.

One significant issue that test developers must address is the question of sex differences (Boyle & Saklofske, 2004). The recent tendency towards production of neutral ("unisex") personality inventories (by removing items that exhibit significant sex differences) makes it well nigh impossible to obtain complete and accurate personality profiles that distinguish between males and females. One example of such an instrument is Morey's (1991) *Personality Assessment Inventory* (PAI), for which a decision was made to minimize sex differences and not only remove items that showed a sex difference, but also not to report gender-based norms. One justification for the decision derives from the nature of the PAI as

a measure of psychopathology where a stronger argument can be made for less emphasis upon the direct assessment of sex differences. Tests of normal personality, however, must address the matter of there being notable sex differences in psychological functioning resulting from differences in genes, brain anatomy, and sex hormone levels, in addition to significant differences in acculturation and social conditioning in some way. The most common method is the provision of separate norms for males and females. Other instruments, such as the *Comrey Personality Scales*, include scales that specifically reflect behavioral and attitudinal differences between males and females. Issues such as sex differences become interwoven with issues such as the prevalence of forms of psychopathology when we consider that domain of content. Several measures that assess psychopathology use the word "personality" in their title in order to reduce the negative reaction among respondents. But at the same time, similar contrasts within the psychopathological domain between the approaches to the development of measures can be seen as with measures of normal personality.

In relation to the assessment of psychopathology, and prior to the release of the PAI instrument, the factor-analytically constructed *Clinical Analysis Questionnaire* (CAQ, Krug, 1980) was developed to provide a Q-data measure of abnormal personality traits. Part 1 of the CAQ measures the traditional 16PF normal personality trait factors, while Part 2 of the instrument provides measures of 12 abnormal personality trait dimensions. Part 2 of the CAQ was constructed from an extensive series of factor analyses that included the entire *Minnesota Multiphasic Personality Inventory* (MMPI) item pool, together with many additional items pertaining to depression and other aspects of psychopathology with the

aim of measuring more fundamental, underlying source-trait dimensions (Boyle, 1990, 2006; Boyle & Comer, 1990; Smith, 1988). Note again the emphasis placed upon the use of factor analysis to empirically derive scales to reflect areas of content that are presumed to be present in the initial pool of items. Recently the CAQ has been upgraded to the *PsychEval Personality Questionnaire* (PEPQ) produced by the *Institute for Personality and Ability Testing* (IPAT). Factor analysis was more prominent in constructing the *Basic Personality Inventory* (BPI, Jackson, 1989). In this case, content dimensions were identified through a factor analysis of the scales of the MMPI and Jackson's *Differential Personality Inventory* (DPI, Jackson & Messick, 1971), which was based upon considerations of symptoms that reflected established domains of psychopathology. This preliminary analysis that was intended to define the domains common to the two measures led to the identification of psychopathological constructs for which a new item pool could be written, with the use of sequential item analytic strategies to finalize the scales.

While the MMPI remains the most widely used measure of psychopathology, indeed of personality in general (Piotrowski & Keller, 1989), the 1989 revision failed to correct significant weaknesses (Helmes & Reddon, 1989). Some of those weaknesses derive directly from the contrasted groups method used to select items for the MMPI scales. Such problems are not as striking with either the CAQ or the DPI and BPI. However, the methods of scale construction used with these measures clearly do influence the contents. For example, the CAQ has scales for *Boredom and Withdrawal, Guilt and Resentment*, *Low Energy Depression*, *Anxious Depression*, and *Suicidal Depression*. The BPI merges into a single *Depression* scale the fine distinctions based upon severity

and associated symptoms that are made by the factor-analytic method used in the

development of the CAQ. Nevertheless, the two instruments converge on scales for the

constructs of hypochondriasis, paranoia/ideas of persecution, anxiety, and thinking

disturbance/schizophrenia. The CAQ has additional scales for *Agitation, Psychopathic*

Deviation, Psychasthenia, and *Psychological Inadequacy*, scales with clear links to the

parent MMPI item pool and its associated psychiatric syndromes, some of which are no

longer used diagnostically. In contrast, the BPI has scales that diverge to assess a wider

range of other forms of psychopathology: *Interpersonal Problems, Alienation, Impulse*

Expression, Social Introversion, and *Self Depreciation*.

Both the *16PF* and the *PRF* have withstood the test of critical scrutiny over many years,

both in the *Test Critiques* series of psychological test reviews, and in the *Buros Mental*

Measurements Yearbooks. Unlike the fifth edition of the 16PF (16PF5; see Conn &

Rieke 1994), the fourth edition with its multiple parallel forms (A, B, C, D, E) has the

decided advantage of being able to attain virtually any desired level of reliability through

the administration of more items. It is important to note that Cattell had repeatedly

advised that at least two forms of the 16 PF (Forms A + B or C + D) should be

administered together, in order to ensure high reliabilities for each of the *16PF* subscales.

In contrast, while parallel forms for the PRF are available for both shorter (A + B), and

longer versions (AA + BB), the current version (PRF-E) is only available in one form.

Differences with the 16PF are evident in other ways, with all PRF forms having

readability evaluated during the item selection stage, while the 16PF uses different forms

for different educational levels: Forms A and B are suitable for use with most adults

whereas Forms C and D are less demanding of vocabulary and administration time, while

Forms E and F are intended for individuals with low literacy levels. While the longer

PRF scales tend to be more reliable because of their increased length, all forms in

themselves are sufficiently reliable for most uses.

Both the PRF and the original forms of the 16PF were developed before the popularity

of the so-called Big Five or Five Factor Model (FFM) that emphasizes five supra-

ordinate dimensions of *Neuroticism, Extraversion,* Openness to Experience*,*

Agreeableness, and *Conscientiousness* (see Goldberg, 1990; Costa & McCrae, 1992;

McCrae & Costa, 1987). Interestingly, both Cattell (1995, Cattell, Boyle, & Chant,

2002) and Jackson (Jackson, Ashton, & Tomes, 1996) are among those who have argued

for additional dimensions, in addition to those who have criticized the Five Factor Model

on other grounds (e.g., Block, 1995). An additional point in debates over the utility of

the FFM is that fewer relevant predictors necessarily account for less variance than does

prediction based on a larger set of primary factors (Mershon & Gorsuch, 1988).

When the consequences of personality assessment might be negative (e.g., admission to a

mental institution; incarceration in a prison), or positive (e.g., job selection; approval from the

therapist; or release from a mental institution) there may be strong motivation (either

conscious or unconscious) to distort responses to personality questionnaire items. In

order to control for motivational/response distortion, many instruments incorporate

various validity and correction scales, ranging from simple "lie scales" such as in the

Eysenck Personality Questionnaire (EPQ-R; Eysenck & Eysenck, 1975) or the L scale of the

MMPI/MMPI-2, social desirability measures, scales for detecting random responses, and

scales that identify other response sets. Measures of psychopathology such as the

MMPI/MMPI-2 (Butcher et al., 1989) frequently have multiple measures to assess the

validity of the responses -- see Bagby et al. (2006) and Helmes (in press), for reviews of the

MMPI-2 validity measures and related issues. The issue of response distortion is complex,

with trait view corrections (Cattell & Krug, 1971) suggesting that there are various distinct

"desirability response tendencies" which differentially distort responses on subjective Q-data

instruments, which can be manipulated all too easily because of their item transparency.

A recent development in the assessment of people where distortion of self-reports or

minimization of reporting problems is likely to be present is based upon conditional

reasoning, a process in which response alternatives are designed to elicit responses

based upon self-serving cognitive premises (James, 1998). This is an indirect or

implicit measure of personality, one that may be more resistant to faking than other,

more traditional methods (LeBreton, Barksdale, Robin, & James, 2007), but which

uses a conventional self-report format. This approach relies upon knowledge of the

forms of self-protective or biased reasoning likely to be used, together with a careful

selection of response alternatives. It is thus quite distinct from measures used for the

implicit assessment of stereotyped attitudes, such as the *Implicit Association Test*

(Greenwald & Banaji, 1995). To-date, relatively few applications of conditional

reasoning have been published, so it is still unclear how extensive the application of

the procedure will be. James et al. (2005) outlined one of the early applications,

namely, the measurement of aggression.

While subjective self-report assessments certainly dominate personality assessment currently, other approaches have definite advantages and continue to be used. Ratings (L-data) based upon previous periods of observation and acquired knowledge of the person being rated have been in use for decades. This form of assessment is particularly useful when there are grounds for believing that self-reports may not be accurate. In addition, such forms of assessment also provide a different source of information that can be useful to the practitioner in developing a more complete understanding of a client. Psychiatric settings are certainly ones in which there are good reasons for doubting the accuracy of many self-reports and instruments for these purposes have been in existence for some time. One of the first examples of a rating scale for psychiatric problems was that developed by Wittenborn (1951). There is now an extensive literature on psychiatric rating scales that cannot be explored here, but Sajatovic and Ramirez (2003) provide examples of many such scales.

A similar context in which there are solid grounds for doubting the accuracy of self-report is the assessment of young children. This has become one of the major areas in which such rating instruments are used, the evaluation of psychosocial functioning and behaviour problems in children (e.g., Achenbach & McConaughy, 2003; Connors, 1997; Reynolds & Kamphaus, 2004). Hartup and van Lieshout (1995) reviewed many of the issues relevant to the assessment of personality during the

course of child development.

The instrument developed by Costa and McCrae (1992) to assess the domains of the Five Factor Model, the NEO-PI-R, incorporates an unusual form of observer rating. The standard self-report form is converted to third-person format. For example, item 1 on Form S, "I am not a worrier", is changed on the male version of Form R to "He is not a worrier" and on the female form to "She is not a worrier". The observer-report Forms R are intended to be completed by a spouse, a peer, or by an expert who knows the individual well. Thus instead of providing ratings for traits based upon provided definitions, or a series of adjectives, Form R of the NEO-PI-R asks the rater to complete the equivalents of the self-report items. Costa and McCrae (1992) reported substantial agreement using intraclass correlations for both the five major domains of the NEO-PI-R, but also for the facet scales across peer/peer, peer/self, and spouse/self comparisons. Such results suggest that the reduction in method variance associated with comparisons of other forms of peer rating with self-reports may lead to better reliability and better agreement on personality characteristics between observers, as noted by Kurtz, Lee, and Sherker (1999). A series of studies of personality change in people with Alzheimer's disease, as rated by spouses and other caregivers, has also shown the utility of the observer-rating forms of the NEO-PI-R (Siegler, Dawson, & Welsh, 1994; Strauss, Pasupathi, & Chatterjee, 1993, Strauss & Pasupathi, 1994; Welleford, Harkins, & Taylor, 1995).

Self-descriptive adjectives have had an extensive history of use in psychological

assessment, as much with experimental methods as with use in applied settings. The

original compilation by Allport and Odbert (1936) has led to the widespread use of

adjectives in various formats and many empirical studies, but with relatively few

widely recognized and used standardized versions. One of the best-known such

measures based on adjectives is the *Multiple Affect Adjective Check List* (MAACL,

Zuckerman & Lubin, 1965), and its subsequent revisions (MAACL-R, Zuckerman &

Lubin, 1985, 1999). A bibliography of research using the MAACL identified over

1900 articles and dissertations (Lubin, Swearrngin, & Zuckerman, 1997). The

original version contained 132 adjectives, of which only 66 were scored for the

domains of Anxiety, Depression, and Hostility. Ratings of the adjectives could be

performed to assess either immediate, "State "responses or longer term, "Trait"

attributes through changes in instructions. The revised versions followed new

analyses with new samples of respondents, and added two scales for Positive Affect

and Sensation Seeking. Of note is that the development of scales for adjective

checklists generally relies upon exploratory factor-analytic approaches, and so these

instruments also become involved in the debates over the most appropriate

methodology that are prominent within the domain of personality assessment. The

MAACL-R exhibits the advantages of adjective checklists, in that it is very flexible in

use and takes little time to complete in comparison with standard multiscale

personality inventories.

A more recent development in the use of adjective ratings for personality assessment

is based upon theories of interpersonal relationships such as those of Benjamin (1974)

and Wiggins (1979). Substantial bodies of theory for both normal personality and

psychopathology have now been developed based on interpersonal models (for

example, Kiesler, 1996; Horowitz, 2004). Such interpersonal models generally

involve dimensions of agency (dominance) and communion (warmth) and measures

are derived on the basis of two-dimensional factor-analytic procedures. Various

specialized instruments have been developed to explore particular interpersonal

models. One that has been established based entirely upon adjectives is the

Interpersonal Adjective Scales (IAS, Wiggins, 1995). The development of the IAS

can be traced to the same Allport and Odbert (1936) list of adjectives that had been

successively refined by Norman (1967). A total of 567 interpersonal adjectives were

classed into 16 categories and successively analyzed and refined to form eight 16-

item scales. The 128 items assess the octant domains of *Assured-Dominant,*

Arrogant-Calculating, Cold-hearted, Aloof-Introverted, Unassured-Submissive,

Unassuming-Ingenuous, Warm-Agreeable, and Gregarious-Extraverted. The

influence of interpersonal theories is likely greater in research than in day to day

psychological assessment practices, but these circumplex models that are used in

interpersonal theories are easily understood and appealing to many psychologists.

The growing body of research in the area should lead to more applications in

professional practice in the future.

There is currently a plethora of "personality tests" and the number has literally exploded in

recent years in both the research literature and from commercial publishers (see the *Buros*

Mental Measurement Yearbooks for the latter). Yet, virtually all of these personality

instruments are subjective self-rating scales or questionnaires, or observer rating scales

(Boyle, 2007). Aside from response sets, and (superficial) conscious reporting, a major

problem with rating scales is that they depend upon transparent, face-valid items unless

extensive developmental work has been done to minimize the influence of irrelevant

processes and to ensure the items are readily understood by the vast majority of respondents.

Otherwise, item transparency may be associated with problematic or invalid responses or be

influenced by motivational distortion (Boyle, in press). The consequence of insufficient

attention to item characteristics is that many current personality assessment instruments may

be based on methodology that can be easily criticized and dismissed, with resulting serious

questions as to the validity of the measure for many purposes. Correction scales can go only

so far (and in some cases, such as that of the MMPI/MMPI-2 *K* correction scale, the

application of such corrections may themselves be problematic). Whereas self-report (Q-

data) personality assessment is based on subjective answers to questions, what is needed is

increased sensitivity to the characteristics of personality items, increased empirical analysis of

items prior to their inclusion in the final version of scales, and the application of the resulting

scales across multiple samples of individuals in order to ensure the generalizability of the

results. Such considerations should be applied to instruments whether they are traditional or

are administered via interactive, computer-based measures of personality. We do know much

about how to ask questions of people about both innocuous and sensitive matters. Awareness

of this material is often not evident among descriptions of personality measures, but more

attention to the constituent items of personality measures and how those instruments were

developed will lead to better measures, evidence-based assessment procedures (Hunsley &

Mash, 2005), and one hopes, to better practices in personality assessment in general.

References

Achenbach, T. M., & McConaughy, S. H. (2003). The Achenbach System of empirically based assessment. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Personality, behavior, and context* (pp. 406-430). New York: Guilford.

Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycho-lexical study. *Psychological Monographs*, 47 (211).

Angleitner, A., John, O. P., & Lohr, F.-J. (1986). It's *what* you ask and *how* you ask it: An itemmetric analysis of personality questionnaires. In A. Angleitner & J. S. Wiggins (Eds.), Personality assessment via questionnaires: Current issues in theory and measurement (pp. 61-108). Berlin: Springer-Verlag.

Bagby, R. M., Marshall, M. B., Bury, A. S., Bacchiochi, J. R., & Miller, L. S. (2006). Assessing underreporting and overreporting response styles on the MMPI-2. In J. N. Butcher (Ed.). MMPI-2: A practitioner's guide (pp. 39-69). Washington, DC: American Psychological Association.

Benjamin, L. S. (1974). Structural analysis of social behavior. *Psychological Review, 81,* 392-425.

Bentler, P. M. (1988). Causal modeling via structural equation systems. In J. R. Nesselroade & R. B. Cattell (Eds.), Handbook of multivariate experimental psychology (2nd ed.). New York: Plenum.

Block, J. (1965). The challenge of response sets: Unconfounding meaning, acquiescence, and social desirability in the MMPI. New York: Appleton-Century-Crofts.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. Psychological Bulletin, 117, 187-215.

Boyle, G. J. (1990). 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. (2006). Scientific analysis of personality and individual differences. Doctor of Science Thesis, St. Lucia, Queensland: University of Queensland.

Boyle, G. J. (2007). An overview of contemporary personality assessment. Paper presented at the International Military Testing Association Conference, Gold Coast, Queensland, October 8-12.

Boyle, G. J. (in press). Personality Questionnaires and Rating Scales – A flawed methodology? In D. Westen, L. Burton, & R. Kowalski (Eds.), Psychology: Australian and New Zealand 2nd edition. Milton, Queensland: Wiley.

Boyle, G. J., & Comer, P. G. (1990). Personality characteristics of direct-service personnel in community residential units. Australia and New Zealand Journal of Developmental Disabilities, 16, 125-131.

Boyle, G. J., & Saklofske, D. H. (2004). (Eds.), Sage benchmarks in psychology: The psychology of individual differences (Vols. 1-4). London: Sage.

Boyle, G. J., Stankov, L., & Cattell, R. B. (1995). Measurement and statistical models in the study of personality and intelligence. In D. H. Saklofske & M. Zeidner (Eds.), International handbook of personality and intelligence (pp. 417-446). New York: Plenum.

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.

Burdsal, C. A., & Vaughn, D. S. (1974). A contrast of the personality structure of college students found in the questionnaire medium by items as compared to parcels. Journal of Genetic Psychology, 135, 219-224.

Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. American Psychologist, 39, 214-227.

Burisch, M. (1986). Methods of personality inventory development – A comparative analysis. In A. Angleitner & J. S. Wiggins (Eds.), Personality assessment via questionnaires: Current issues in theory and measurement (pp. 109-120). Berlin: Springer-Verlag.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). Minnesota Multiphasic Personality Inventory-2. Manual for administration and scoring. Minneapolis, MN: University of Minnesota Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

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

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

Cattell, R. B. (1986a). General principles across the media of assessment. In R. B. Cattell & R. C. Johnson (Eds.), Functional psychological testing: Principles and instruments (pp. 15-32). New York:Brunner/Mazel.

Cattell, R. B. (1986b). Selecting, administering, scoring, recording, and using tests in

assessment. In R. B. Cattell & R. C. Johnson (Eds.), (1986). <u>Functional psychological</u>

<u>testing: Principles and instruments</u> (pp. 105-126). New York: Brunner/Mazel.

Cattell, R. B. (1986c). Structured tests and functional diagnoses. In R. B. Cattell & R. C.

Johnson (Eds.), <u>Functional psychological testing: Principles and instruments</u> (pp. 3-

14). New York: Brunner/Mazel.

Cattell, R. B. (1986d).The psychometric properties of tests: Consistency, validity, and

efficiency. In R. B. Cattell & R. C. Johnson (Eds.), <u>Functional psychological testing:</u>

<u>Principles and instruments</u> (pp. 54-78). New York: Brunner/Mazel.

Cattell, R. B. (1988). The meaning and strategic use of factor analysis. In J. R.

Nesselroade & R. B. Cattell (Eds.), <u>Handbook of multivariate experimental</u>

<u>psychology</u> (2nd ed.). New York: Plenum.

Cattell, R. B. (1994). A cross-validation of primary personality structure in the 16 P.F. by

two parcelled factor analyses. <u>Multivariate Experimental Clinical Research, 10</u>, 181-

190.

Cattell, R. B. (1995). The fallacy of five factors in the personality sphere. <u>The Psychologist,</u>

<u>May</u>, 207-208.

Cattell, R. B., Boyle, G. J., & Chant, D. (2002). The enriched behavioral prediction

equation and its impact on structured learning and the dynamic calculus.

<u>Psychological Review, 109</u>, 202-205.

Cattell, R. B., Cattell, A. K., & Cattell, H. E. P. (1994). *The Sixteen Personality Factor*

Questionnaire (5th ed.). Champaign, IL: Institute for Personality and Ability

Testing.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). Handbook for the Sixteen Personality Factor Questionnaire (16PF). Champaign, IL: Institute for Personality and Ability Testing.

Cattell, R. B., & Krug, S. E. (1971). A test of the trait-view theory of distortion in measurement of personality questionnaire. Educational and Psychological Measurement, 31, 721-734.

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 Objective-Analytic (O-A) Test Kit. Champaign, IL: Institute for Personality and Ability Testing.

Comrey, A. L. (1967). Tandem criteria for analytic rotation in factor analysis. *Psychometrika, 32*, 143-154.

Comrey, A. L. (1984). Comparison of two methods to identify major personality factors. *Applied Psychological Measurement, 8*, 397-408.

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. Journal of Consulting and Clinical Psychology, 56, 754-761.

Conn, S. R., & Rieke, M. L. (1994). Technical Manual for the 16 P.F.(5th ed.) Champaign, IL, Institute for Personality and Ability Testing.

Conners, C. K. (1997). *Conners' Rating Scales – Revised technical manual*. North Tonawanda, NY: Multi-Health Systems Inc.

Costa, P. T., Jr., & McCrae, R. R. (1992). Revised NEO Personality Inventory and NEO Five-Factor Inventory: Professional manual. Odessa, FL: Psychological Assessment Resources.

Cronbach, L., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-303.

Cuttance, P. (1987). Issues and problems in the application of structural equation models. In P. Cuttance, & R. Ecob (Eds.), Structural modeling by example: Applications in educational, sociological, and behavioral research (pp. 241-279). Cambridge, UK: Cambridge University Press.

Drasgow, F., & Olson-Buchanan, J. B. (Eds.).(1999). Innovations in computerized assessment. Mahwah, NJ: Erlbaum.

Edwards, A. L. (1957). The social desirability variable in personality assessment and research. New York: Dryden.

Eysenck, H. J., & Eysenck, S. B. G. (1975). Manual of the Eysenck Personality Questionnaire (junior and adult). London: Hodder & Stoughton.

Goffin, R. D., Rothstein, M. G., & Johnston, N. G. (2000). Predicting job performance using personality constructs: Are personality tests created equal? In R. D. Goffin & E. Helmes (Eds.), Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy (pp. 249-264). New York: Kluwer.

Goldberg, L. R. (1990). An alternative "description of personality": The Big Five factor structure. Journal of Personality and Social Psychology, 59, 1216-1229.

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

Gorsuch, R. L. (1988). Exploratory factor analysis. In J. R. Nesselroade & R. B. Cattell (Eds.), Handbook of multivariate experimental psychology (2nd ed.). New York: Plenum.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. Psychological Review, 102, 4-27.

Groth-Marnat, G. (2003). Handbook of psychological assessment (4th ed.). Hoboken, NJ: Wiley.

Haggbloom, S. J., Warnick, R., Warnick, J. E., Jones, V. K., Yarbrough, G. L., Russell, T. M., Borecky, C. M., McGahhey, R., Powell III, J. L., Beavers, J., & Monte, E. (2002). The 100 most eminent psychologists of the 20th century. Review of General Psychology, 6, 139-152.

Hartup, W. W., & van Lieshout, C. F. M. (1995). Personality development in social context. Annual Review of Psychology, 46, 655-687.

Helmes, E. (2000). The role of social desirability in the assessment of personality constructs. In R. D. Goffin & E. Helmes (Eds.), Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy (pp. 21-40). New York: Kluwer.

Helmes, E. (in press). Response distortion in applications of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) in offender rehabilitation. Journal of Offender Rehabilitation.

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.

Horowitz, L. M. (2004). *Interpersonal foundations of psychopathology*. Washington, DC: American Psychological Association.

Hough, L., & Paullin, C. (1994). Construct-oriented scale construction: The rational approach. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.). Biodata handbook: Theory, research, and used of biographical information in selection and performance (pp. 109-145). Palo Alto, CA: Consulting Psychologists Press.

Hunsley, J., Lee, C. M., & Wood, J. M. (2003). Controversial and questionable assessment techniques. In S. O. Lilienfeld, Lynn, S. J., & J. M. Lohr (Eds.). Science and pseudoscience in clinical psychology (pp. 17-38). New York: Guilford.

Hunsley, J. & Mash, E. J. (2005). Introduction to the Special Section on Developing Guidelines for the Evidence-Based Assessment (EBA) of adult disorders. . Psychological Assessment, 17, 251-255.

Jackson, D. N. (1970). A sequential system for personality scale construction. In C. D. Spielberger (Ed.), Current topic in clinical and community psychology. (Vol. 2, pp. 61-96). New York: Academic Press.

Jackson, D. N. (1984). Personality Research Form manual. Port Huron, MI: Research Psychologists Press.

Jackson, D. N. (1989). Basic Personality Inventory manual. Port Huron, MI: Sigma Assessment Systems.

Jackson, D. N. (1994). Jackson Personality Inventory – Revised manual. Port Huron, MI & London, Ontario: Sigma Assessment Systems.

Jackson, D. N. (2000). A perspective. In R. D. Goffin & E. Helmes (Eds.). <u>Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy</u> (pp. 333-344). New York: Kluwer Academic.

Jackson, D. N., Ashton, M. C., & Tomes, J. L. (1996). The six-factor model of personality: Facets from the big five. <u>Personality and Individual Differences, 21,U 391-402.</u>

Jackson, D. N. & Messick, S. (1971). *The Differential Personality Inventory*. London, Ontario: Authors.

James, L. R. (1998). Measurement of personality via conditional reasoning. <u>Organizational Research Methods, 1,</u> 131-163.

James, L. R., McIntyre, M. D., Glisson, C. A., Green, P. D., Patton, T. W., LeBreton, J. M., Frost, B. C., Russell, S. M., Sablynski, C. J., Mitchell, T. R., & Williams, L. J. (2005). A conditional reasoning measure for aggression. <u>Organizational Research Methods, 8,</u> 69-99.

Kiesler, D. J. (1996). *Contemporary interpersonal theory and research: Personality, psychopathology and psychotherapy*. New York: Wiley.

Krug, S. E. (1980). <u>Clinical Analysis Questionnaire Manual</u>. Champaign, IL: Institute for Personality and Ability Testing.

Krug, S. E. (1981). <u>Interpreting 16PF profile patterns</u>. Champaign, IL, Institute for Personality and Ability Testing.

Kurtz, I. C., Dawson, D. V., & Welsh, K. A. (1994). Caregiver ratings of personality change in Alzheimer's disease patients: A replication. <u>Psychology and Aging, 9,</u> 464-466.

Kurtz, J. E., Lee, P. A., & Sherker, J. L. (1999). Internal and temporal reliability estimates for informant ratings of personality using the NEO-PI-R and IAS. Assessment, 6, 103-114.

LeBreton, J. M., Barksdale, C. D., Robin, J., & James, L. R. (2007). Measurement issues associated with conditional reasoning tests: Indirect measurement and test faking. Journal of Applied Psychology, 92, 1-16.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. Psychological Reports, 3, 635-694.

Lubin, B., Swearngin, S. E., & Zuckerman, M. (1997). Research with the Multiple Affect Adjective Check List (MAACL and the MAACL-R: 1960-1996). San Diego, CA: Educational and Industrial testing Service.

MacCallum, R. (1985). Some problems in the process of model modification in covariance structure modeling. Paper presented at the European Meeting of the Psychonomic Society, Cambridge, UK.

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

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

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.

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

Mount, M. K., Witt, L. A., & Barrick, M. R. (2000). Incremental validity of empirically keyed biodata scales over GMA and the five factor personality constructs. Personnel Psychology, 53, 299-323.

Mulaik, S. A. (1988). Confirmatory factor analysis. In J. R. Nesselroade & R. B. Cattell (Eds.), Handbook of multivariate experimental psychology (2nd ed.). New York: Plenum.

Nichols, J. G., Licht, B. G., & Pearl, R. A. (1982). Some dangers of using personality questionnaires to study personality. Psychological Bulletin, 92, 572-580.

Norman, W. T. (1967). *2800 personality trait descriptors: Normative operating characteristics for a university population.* Department of Psychology, University of Michigan, Ann Arbor.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. Journal of Personality and Social Psychology, 44, 598-609.

Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleitner & J. S. Wiggins (Eds.). Personality assessment via questionnaires: Current issues in theory and measurement (pp. 143-165). Berlin: Springer-Verlag

Paulhus, D. L. (1998). Paulhus Deception Scales (PDS): The Balanced Inventory of Desirable Responding-7. User's manual. Toronto: Multi-Health Systems.

Piotrowski, C. & Keller, J. W. (1989). Psychological testing in outpatient mental health facilities: A national study. Professional Psychology: Research and Practice, 20, 423-425.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children, Second Edition manual*. Circles Pines, MN: AGS.

Rudinger, G. & Dommel, N. (1986). An example of convergent and discriminant validation of personality questionnaires. In A. Angleitner & J. S. Wiggins (Eds.). Personality assessment via questionnaires: Current issues in theory and measurement (pp. 214-224). Berlin: Springer-Verlag.

Sajatovic, M. & Ramirez, L. F. (2003). Rating scales in mental health (2nd ed.). Hudson, OH: Lexi-Comp.

Schuerger, J. M. (1986). Personality assessment by objective tests. In R. B. Cattell & R. C. Johnson (Eds.), Functional psychological testing: Principles and instruments (pp. 260-287). New York: Brunner/Mazel.

Schwarz, N. (1999). Frequency reports of physical symptoms and health behaviors: How the questionnaire determines the results. In D. C. Park, R. W. Morrell, & K. Shifrin (Eds.), Processing of medical information in aging patients: Cognitive and human factors perspectives (pp. 93-108). Mahwah, NJ: Erlbaum.

Schwarz, N. & Sudman, S. (1996). (Eds.), Answering questions: Methodology for determining cognitive and communicative processes in survey research. San Francisco, CA: Jossey-Bass.

Smith, B. D. (1988). Personality: Multivariate systems theory and research. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed.). New York: Plenum.

Stokes, G. S., Mumford, M. D., & Owens, W. A. (1994). Biodata handbook: Theory, research, and used of biographical information in selection and performance. Palo Alto, CA: Consulting Psychologists Press.

Strauss, M. E., Pasupathi, M., & Chatterjee, A. (1993). Concordance between observers in descriptions of personality change in Alzheimer's disease. Psychology and Aging, 8, 475-480.

Strauss, M. E. & Pasupathi, M. (1994). Primary caregivers' descriptions of Alzheimer patients' personality traits: Temporal stability and sensitivity to change. Alzheimer Disease and Associated Disorders, 8, 166-176.

Sudman, S. & Bradburn, N. M. (1982). Asking questions: A practical guide to questionnaire design. San Francisco: Jossey-Bass.

Velicer, W. F. & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25,* 1-28.

Weiner, I. B. (1997). Current status of the Rorschach inkblot method. Journal of Personality Assessment, 68, 5-19.

Welleford, E. A., Harkins, S. W., & Taylor, J. R. (1995). Personality change in dementia of the Alzheimer's type: Relations to caregiver personality and burden. Experimental Aging Research, 21, 295-314.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, *37*, 395-412.

Wiggins, J. S. (1995). *IAS. Interpersonal Adjective Scales: Professional manual.* Odessa, FL: Professional Assessment Resources.

Wirth, R. J. & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. Psychological Methods, 12, 58-79.

Wittenborn, J. R. (1951). Symptom patterns in a group of mental hospital patients.

    Journal of Consulting Psychology, 15, 290-302.

Zuckerman, M. & Lubin, B. (1965). Manual for the Multiple Affect Adjective Check

    List. San Diego, CA: Educational and Industrial Testing Service.

Zuckerman, M. & Lubin, B. (1985). Manual for the Multiple Affect Adjective Check

    List - Revised. San Diego, CA: Educational and Industrial Testing Service.

Zuckerman, M. & Lubin, B. (1999). Manual for the MAACL-R Multiple Affect

    Adjective Check List. 1999 Edition. San Diego, CA: Educational and Industrial

    Testing Service.