CHAPTER XIV

A historical survey of personality scales and inventories

Lewis R. Goldberg

One of the most salient developments in psychological assessment has been the recent proliferation of multi-scale personality inventories. Yet, historical accounts of the testing movement (e.g., Du Bois, 1970) have given short shrift to this important development. The present chapter, which represents an attempt to fill this historical lacuna, will consider the precursors of today’s personality inventories. The contents of this chapter should provide the historical material for later interpretive analyses of the particular traits focused upon by previous investigators (e.g., Goldberg, 1970) and for analyses of the strategies used to measure these traits (see Hase & Goldberg, 1967). The chapter is focused exclusively on structured personality measures; no consideration is given to significant milestones in aptitude or achievement testing, or to such less structured measures as projective techniques (e.g., Downey’s Will Profile) and rating procedures (e.g., the Vineland Social Maturity Scales).

A capsule summary of this historical material is presented in Table 1. The dates listed in Table 1 are only approximate, as the work on a scale or inventory from conception to the first published account may take five years, and to the first published manual up to ten years. Consequently, the dates—which are based upon published reports—are systematically biased, the initial work having been carried out prior to the date listed. Moreover, the entries in Table 1 may not represent a complete collection of historical milestones. First of all, they are overwhelmingly of American origin. Second, with only a few exceptions, all of the milestones either have been published commercially or have been reported in the general psychological literature; consequently, many scales developed for industrial and military organizations have not been included. Third, many of the scales and inventories have been revised over the years, and most of these revisions have not been included in the table. Finally, the entries listed in Table 1 do not include any of the purely methodological books or papers which have profoundly influenced personality scale construction.

The scales and inventories listed in Table 1 have been classified by the type

Financial support during the preparation of this paper was provided by Grant No. MH 12972 from the National Institute of Mental Health, U.S. Public Health Service.
of individual differences they purport to measure. The first three columns in the table include those measures which have been developed as a response to applied societal pressures, namely to forecast (a) personal or social adjustment, (b) satisfaction and success in vocational choice, and (c) academic achievement (above and beyond that predictable from scholastic aptitude tests). The remaining columns in the table list measures which were less directly stimulated by applied demands, including those scales and inventories directed at two extraordinarily popular targets for structured measurement, namely introversion-extroversion and masculinity-femininity, and at two influential theories of individual differences, namely Spranger’s (1928) schema for classifying “personal values” and Murray et al.’s (1938) classification of “manifest needs.”

While these distinctions are extremely useful for understanding the development of the early scales, the proliferation of more comprehensive inventories after World War II has made the categorization of these recent measures a difficult undertaking. For example, the majority of modern personality inventories include scales designed to assess aspects of adjustment, introversion, masculinity, scholastic potential, and personal values and needs, and many of these inventories could therefore have been included in each of the major categories. Consequently, one must consider the above distinctions with some caution, since they inevitably reflect the idiosyncrasies of one individual’s efforts to reconstruct history on the basis of often-incomplete published accounts. Nonetheless, this chapter may represent the first reasonably systematic effort to link together a number of convergent developments in the history of structured measurement techniques, and a careful perusal of Table 1 should lead to some insights regarding the most salient individual differences included for measurement and the strategies employed to assess them.

PSYCHOPATHOLOGY VERSUS ADJUSTMENT

By far the most significant class of individual differences addressed by developers of personality scales and inventories includes those traits of temperament or character which purportedly distinguish the better “adjusted” members of our society from their “neurotic,” “psychotic,” or “criminal” neighbors. The measurement of aspects of emotional or social adjustment (or, not necessarily the converse, aspects of neuroticism, psychoticism, or psychopathy) has a relatively long history in psychology, stemming in large part from societal pressures on psychologists to devise indicators of potential emotional breakdown in military and industrial settings. While many early scholars had discussed aspects of personal and social adjustment, it remained for Freud to popularize the topic through the wide currency given his theories in medical,”

---

1 Throughout this chapter, the names of test developers and the dates of publication of a scale or inventory do not necessarily correspond to the references in the Bibliography at the end of this volume. Rather, where possible, these names and dates are based on the citations in Buros (1970). On the other hand, all dates cited in parentheses—e.g., Buros (1970)—refer to published articles or books, and all of these publications are included in the Bibliography.

2 For the reader’s convenience, a list of the abbreviations used in this chapter is included in Table 1.
<table>
<thead>
<tr>
<th>Approx. Date</th>
<th>Psychopathology-Adjustment</th>
<th>Vocational Interests</th>
<th>Scholastic Predictors</th>
<th>Introversion-Extroversion</th>
<th>Masculinity-Femininity</th>
<th>Personal Values &amp; Manifest Needs</th>
<th>Other Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1906</td>
<td>Neyman &amp; Warrane (Developed symptom list)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1913</td>
<td>Hoch &amp; Amsden (Revised list)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1914</td>
<td>Wells (Revised list)</td>
<td>Kelley (Early interest questionnaire)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1916</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Jung (Proposed theory)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1917</td>
<td>Woodworth-Personal Data Sheet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1919</td>
<td></td>
<td>Yoakum et al. (Developed item pool)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1920</td>
<td>Johnson (Revised PDS for children)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Alport &amp; Allport (Theoretical paper)</td>
<td></td>
</tr>
<tr>
<td>1921</td>
<td>Pressey-X-O Tests for Investigating the Emotions</td>
<td>Moore-Sales vs. Design Engineering Score</td>
<td>Pressey-X-O Tests</td>
<td>McDougall (Extended theory)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1922</td>
<td></td>
<td>Freyd-Two interest scales for engineers</td>
<td>Miner-Interest scales for high school students</td>
<td>Terman (Began studies of male vs. female reactions and preferences)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1923</td>
<td>Mathews (Revised PDS for children)</td>
<td>Cady (Revised PDS for juvenile delinquents)</td>
<td>Conklin (Revised I-E theories)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1924</td>
<td></td>
<td>Ream-Interests of successful vs. unsuccessful life insurance salesmen</td>
<td>Freyd (Developed list of 54 I-E traits)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1925</td>
<td>Laird-Colgate Mental Hygiene Inventory-Form B</td>
<td>Cowdrey-Interests of physicians, lawyers, and engineers</td>
<td>Laird-Colgate Mental Hygiene Inventory-Form C</td>
<td></td>
<td>Travis-Diagnostic Character Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1926</td>
<td></td>
<td>Strong-Interest scale for personnel managers</td>
<td></td>
<td></td>
<td>Haldbreder-I-E Scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1927</td>
<td>House-Mental Hygiene Inventory</td>
<td>Strong-Vocational Interest Blank</td>
<td></td>
<td></td>
<td>Conklin-I-E Interest Ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1928</td>
<td></td>
<td></td>
<td>Allport &amp; Allport-Ascendance-Submission Reaction Study</td>
<td></td>
<td>Spranger (Theory translated into English)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1929</td>
<td></td>
<td></td>
<td>Neymann &amp; Kohantel-New Diagnostic Test for I-E Scale</td>
<td></td>
<td>Whitman-Short I-E scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1930</td>
<td>Thurstone-Thurstone-Personality Schedule, Jasper-Depression-Elation Scale, Symonds &amp; Jackson-Adjustment Survey</td>
<td>Garrettson &amp; Symonds-Interest Questionnaire for High School Students</td>
<td></td>
<td></td>
<td>Root-I-E Inventory</td>
<td>Vernon &amp; Allport-Study of Values</td>
<td></td>
</tr>
</tbody>
</table>

*Table 1: Some Milestones in the History of Personality Scales and Inventories (See page 303 for key to abbreviations).*
<table>
<thead>
<tr>
<th>Year</th>
<th>Psychopathology-Adjustment</th>
<th>Vocational Interests</th>
<th>Scholastic Predictors</th>
<th>Introversion-Extroversion</th>
<th>Masculinity-Femininity</th>
<th>Personal Values &amp; Manifest Needs</th>
<th>Other Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1932</td>
<td>Willoughby—Emotional maturity scale</td>
<td>Brainard &amp; Stewart—Specific Interest Inventory</td>
<td>Bernoulli—Self-sufficiency scale</td>
<td>Bernoulli—Personality Inventory</td>
<td>Carte—M-F scale from SVIB</td>
<td>Wang—Persistence Scale</td>
<td></td>
</tr>
<tr>
<td>1934</td>
<td>Guilford—NPI Bell—Adjustment Inventory</td>
<td>Kuder—Preference Record—Vocational Interest Schedule</td>
<td>Wechsler—Study Habits Inventory</td>
<td>Guilford—Nebraska Personality Inventory</td>
<td>Guilford—NPI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1935</td>
<td>Humm &amp; Wadsworth—Temperament Schedule Washburne—Social Adjustment Inventory</td>
<td>Thurstone—Vocational Interest Schedule</td>
<td>Flanagan (Factor scales from Bernoulli’s Personality Inventory)</td>
<td>Lente et al.—Conservatism—Radicalism Opinionnaire</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1936</td>
<td>Rundquist &amp; Sletto—Minnesota Scale for the Survey of Opinions</td>
<td>Clesey—Vocational Interest Inventory</td>
<td>Williamson &amp; Darley—Minnesota Inventory of Social Attitudes</td>
<td>Terman &amp; Miles—Attitude-Interest Analysis Test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1937</td>
<td></td>
<td>Dunlap—Academic Preference Blank</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1938</td>
<td>McFarland &amp; Sett—PS Experience Blank</td>
<td>Strong—Vocational Interest Blank (Revised)</td>
<td>Strong—SVIB (Revised)</td>
<td>Murray et al. (Developed list of needs, and initial item pools)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1939</td>
<td>Tiesg, Clark, &amp; Thorpe—California Test of Personality</td>
<td>Gentry—Vocational Inventory</td>
<td>Guilford &amp; Guilford—Inventory of Factors STDCR</td>
<td>Watson &amp; Fisher—Inventory of Affective Potency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1940</td>
<td>Adams &amp; Lepley—Personal Audit Darley &amp; McNamara—Minnesota Personality Scale Fisher &amp; Witson—Inventory of Affective Tolerance Johnson—Temperament Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1942</td>
<td></td>
<td>Cardall—Primary Business Interests Test</td>
<td></td>
<td>Evans &amp; McCon-Neil—Minnesota F-S-E Inventory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1943</td>
<td>Hathaway &amp; McKinley—Minnesota Multifaceted Personality Inventory</td>
<td>Kohl et al.—Inventory of Vocational Interests</td>
<td>Guilford &amp; Martin—GAMIN</td>
<td>Hathaway &amp; McKinley—MMPI (M)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1944</td>
<td>Weider et al.—Cornell Index</td>
<td></td>
<td></td>
<td>Guilford &amp; Martin—Inventory of Factors GAMIN &amp; Personnel Inventory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1945</td>
<td>Maslow et al.—Security-Insecurity Inventory</td>
<td>Van Allyn—Job Qualification Inventory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1946</td>
<td>Thorpe &amp; Clark—Mental Health Analysis</td>
<td>Gregory—Academic Interest Inventory</td>
<td>Drake—MMPI (SI)</td>
<td>Kuder—PR—V (Revised)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gray &amp; Wheelwright—Psychological Type Questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Psychopathology-Adjustment</td>
<td>Vocational Interests</td>
<td>Scholastic Predictors</td>
<td>Introversion-Extroversion</td>
<td>Masculinity-Femininity</td>
<td>Personal Values &amp; Manifest Needs</td>
<td>Other Traits</td>
</tr>
<tr>
<td>------</td>
<td>---------------------------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>--------------------------</td>
<td>------------------------</td>
<td>---------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>1947</td>
<td>Baldwin-Motivation Indicator</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bell-Personal Preference Inventory</td>
<td></td>
</tr>
<tr>
<td>1948</td>
<td>Guilford, Schniedman, &amp; Zimmerman-Interest Survey</td>
<td>Kuder-Preference Record-Personal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1949</td>
<td>Heston-Personal Adjustment Inventory</td>
<td>Brown-Collage Inventory of Academic Adjustment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1951</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1952</td>
<td>Mandler &amp; Sarsen-Test Anxiety Questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1953</td>
<td>Becht &amp; Layton-Minnesota Counseling Inventory</td>
<td>Brown &amp; Holtzman-Survey of Study Habits and Attitudes</td>
<td></td>
<td></td>
<td></td>
<td>Edwards-Personal Preference Schedule</td>
<td>Gordon-Personal Profile</td>
</tr>
<tr>
<td>1954</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1956</td>
<td>Kuder-Preference Record-Ocuppational</td>
<td>Gough-CPI (Ac, Al, Fe)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1957</td>
<td>Jenkins-How Well Do You Know Your Interests?</td>
<td>Heist &amp; Williams-CPI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1958</td>
<td>Weingarten-Picture Interest Inventory</td>
<td>I. G. Sarason-Test Anxiety Scale</td>
<td>Carter-California Study Methods Survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1959</td>
<td>Eysenck-Maudsley Personality Inventory (Neuroticism)</td>
<td>Geist-Picture Interest Inventory</td>
<td>Carter-Interest Scale</td>
<td></td>
<td></td>
<td>Eysenck-MPI (Extroversion) Welsh-WFPT</td>
<td>Welsh-FIGURE Preference Test</td>
</tr>
<tr>
<td>1960</td>
<td>Alpert &amp; Haber-Achievement Anxiety Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Holland-VPI</td>
<td>Gordon-Survey of Interpersonal Values</td>
</tr>
</tbody>
</table>

**Other Traits:**
- Cattell & Stice-Sixteen Personality Factor Questionnaire
- Guilford & Zimmerman-Temperament Survey
- Thurstone-Temperament Schedule
- Adorno et al.-California F Scale
- Allport-Vernon-Lindzey-Study of Values (Revised)
- LaForge & Secrat-Interpersonal Check List
- Gough-Colorado Psychological Inventory
- Heist & Williams-Omnibus Personality Inventory
- Rotter (Developed Theory)
<table>
<thead>
<tr>
<th>Year</th>
<th>Author(s)</th>
<th>Instrument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951</td>
<td>Schene &amp; Charnov</td>
<td>Personality Assessment Test</td>
<td>40-item inventory of personality traits</td>
</tr>
<tr>
<td>1952</td>
<td>Haggard &amp; Frick</td>
<td>Strong Inventory</td>
<td>144-item inventory of personality traits</td>
</tr>
<tr>
<td>1953</td>
<td>Frick &amp; OAS (B)</td>
<td>OAS (B)</td>
<td>120-item inventory of social desirability</td>
</tr>
<tr>
<td>1954</td>
<td>Eyberg &amp; Eyman</td>
<td>Personality Inventory</td>
<td>150-item inventory of personality traits</td>
</tr>
<tr>
<td>1955</td>
<td>Campbell et al</td>
<td>Strong Vocational Interest Blank</td>
<td>250-item inventory of vocational interests</td>
</tr>
<tr>
<td>1956</td>
<td>Gough &amp; Hurlbut</td>
<td>Adjective Check List</td>
<td>500-item inventory of personality traits</td>
</tr>
<tr>
<td>1957</td>
<td>Conley &amp; CPS</td>
<td>CPS</td>
<td>300-item inventory of personality traits</td>
</tr>
<tr>
<td>1958</td>
<td>Goldstein &amp; Kline</td>
<td>KSAT</td>
<td>100-item inventory of academic interests</td>
</tr>
<tr>
<td>1959</td>
<td>PRF</td>
<td>Personality Research Form</td>
<td>200-item inventory of personality traits</td>
</tr>
</tbody>
</table>

**Key to Abbreviations**
- ACI: Adjective Check List
- ASI: Academic Self-Inventory
- ASI: Adaptive Social Index
- AET: Academic Test of Academic Readiness
- AIA: Academic Interest Assessment
- AOA: Academic Orientation Inventory
- AOA: Academic Orientation Scale
- AOS: Academic Orientation Survey
- BSI: Berkeley Self-Inventory
- CPS: Congruent Personality Scale
- DIT: dimension of temperament inventory
- EPT: Educational Placement Test
- EPQ: Eysenck Personality Questionnaire
- GAT: General Aptitude Test
- GAI: General Aptitude Inventory
- GAF: General Aptitude Factor
- GAI: General Aptitude Index
- GAI: General Aptitude Inventory
- GAT: General Aptitude Test
- GSI: General Self-Inventory
- HCR: Haggard Correlation Test
- HCT: Haggard Correlation Test
- HSS: Haggard Self-Understanding Scale
- ITP: Interpersonal Test
- JPI: Jackson Personality Inventory
- LPI: Lifestyle Inventory
- MCI: Marlowe-Crowne Inventory
- MPQ: Maudsley Personality Questionnaire
- OAS: Osterhout and Abrahams Score
- OIS: Osterhout and Sumner Scale
- PPI: Personality Preference Inventory
- PRF: Personality Research Form
all words associated with behaviors the subject considered “wrong,” and to
circle the worst of these. A fourth form (e.g., injustice—noise—self-conscious-
ness—discouragement—germs) contained instructions for the subjects to cross
out all the things in each list about which they have ever been worried and to
circle the most worrisome. Pressey’s adjustment scales, like those from Travis’s
(1925) Diagnostic Character Test, were constructed by keying statistically
in frequent responses; the implicit assumption of both Pressey and Travis that
social and emotional maladjustment is related to unconventional test response
patterns was later explicitly propounded by Berg (1955) as the “Deviation
Hypothesis.”

More conventional early attempts to measure aspects of psychopathology
included Symonds and Jackson’s (1930) Adjustment Survey, Jasper’s (1930)
Depression-Elation Scale, Willoughby’s (1932) Emotional Maturity Scale, and
Washburne’s (1935) Social Adjustment Inventory. During the same period,
however, these global (i.e., single-scale) measures of adjustment began to be
supplemented by multi-scale inventories, each scale designed to measure some
different facet of psychopathology or adjustment. The forerunner of these
inventories was one developed in 1925 by Travis, who attempted to measure
50 “traits” (e.g., stability, inferiority, narcissism, hypochondria, melancholia,
sadism, paranoia), each by means of two-item “scales.” While inventories
composed of much longer scales were soon used to measure vocational in-
terests (e.g., the first version of Strong’s Vocational Interest Blank, published
in 1927), it remained for Vernon and Allport (1931) and Bernreuter (1933b)
to popularize the construction of more comprehensive personality inventories.

Bernreuter combined the items from Thurstone and Thurstone’s Personality
Schedule, Laird’s Colgate Mental Hygiene Inventory (Form C), Allport
and Allport’s Ascendence-Submission Reaction Study, and his own Self-
Sufficiency Scale into one large pool, which he then administered to some
400 college students. For each of the four scales in turn, Bernreuter selected
the 50 subjects with the highest scores and the 50 subjects with the lowest
scores to be used as criterion groups for an empirical analysis of each item in
the combined pool; the resulting four overlapping sets of differentiating items
comprised his Personality Inventory. Bernreuter’s four scales proved to be at
least moderately reliable and to correlate highly with each of the original
scales they were meant to replace. Unfortunately, they also turned out to be
intercorrelated so highly (e.g., the Neurotic Tendencies and Introversion-
Extraversion scales correlated .96) that Flanagan (1935) could later reproduce
virtually all of their variance with two orthogonal factor scales. By this time,
however, the popular demand for multi-scale inventories was so great that
Bernreuter elected to include all six scales (the four original scales plus the
two factor scales) in the 1938 revision of his inventory.

Meanwhile, Bell attempted to measure four separate aspects of adjustment
with his Adjustment Inventory, which was published in 1934; this popular
inventory yielded scores on Home, Health, Social, and Emotional Adjust-
ment, plus a total (composite) score (see Bell, 1935). A similar inventory,
Rundquist and Sletto’s (1936) Minnesota Scale for the Survey of Opinions,
included scales for General Adjustment, Morale vs. Insecurity, Inferiority vs.

Social Ease, Family Intimacy vs. Discord, Respect vs. Disrespect for the
Law, Economic Conservatism vs. Radicalism, and Respect vs. Contempt for
Education.

An analogous inventory for children, the California Test of Personality, was
published by Tigs, Clark, and Thorpe in 1939; its scales purportedly
assessed various aspects of self-adjustment as Self-Reliance, Sense of Personal
Worth, Sense of Personal Freedom, Feeling of Belonging, Withdrawal Tend-
encies, and Nervous Symptoms; plus such aspects of social adjustment as
Social Standards, Social Skills, Antisocial Tendencies, and Family, School,
and Community Relations. A revision of this inventory was published by
Thorpe and Clark in 1946; called the Mental Health Analysis, this later
inventory was designed to measure five categories of mental health “assets”
(Personal Relationships, Interpersonal Skills, Social Participation, Satisfaction
with Work and Recreation, and Adequacy of Outlook and Goals), plus five
categories of “liabilities” (Immaturity, Instability, Feelings of Inadequacy,
Physical Defects, and Nervous Manifestations), as well as to furnish two part-
scores (Assets and Liabilities) and a total Adjustment Index.

By the mid-1930s some inventory developers began turning to diagnosed
psychiatric patients as criterion groups for the development of adjustment
scales and inventories. This external strategy of scale construction had been
used as early as 1921 by Pressey to develop a scale of scholastic potential and
by Moore to differentiate sales vs. design engineers. In the 1920s a number
of vocational interest scales and inventories utilized the same strategy, the most
famous being Strong’s Vocational Interest Blank in 1927. However, it was not
until the middle 1930s and the development of Humm and Wadsworth’s
(1934, 1935) Temperament Schedule that this strategy was used to construct
adjustment scales. The Temperament Schedule, published in 1935, originally
yielded scores on seven of the components of adjustment vs. psychopathology
defined by the psychiatrist Rosanoff (1920): Normal, Hysteroid, Manic, De-
pressive, Autistic, Paranoid, and Epileptoid. The seven scales were developed
by item analysis against criterion groups of psychiatric patients and normal
subjects who had been judged as being extreme on the component. Interest-
ingly, while the Temperament Schedule was developed from the responses of
psychiatric patients, the inventory appears to have enjoyed far more popularity
in industrial than in clinical settings.

The publication of the Temperament Schedule was soon followed by the
development of another adjustment inventory by the same strategy, namely
the P-S Experience Blank (Psycho-Somatic Inventory) reported by McFarland
and Seitz (1938). The two highly correlated scales in this inventory, designed
to measure physiological dysfunction and psychological maladjustment, were
constructed by contrasting the response of “neurotic” and “normal” subjects.

This same external strategy was used to construct what was to become the
most popular of all adjustment inventories, Hathaway and McKinley’s Minne-
sota Multiphasic Personality Inventory (MMPI), which was developed in the
late 1930s and published in 1943. Each of the original MMPI clinical scales
(Hypochondriasis, Depression, Hysteria, Psychopathic Deviancy, Paranola,
Psychasthenia, Schizophrenia, and Hypomania) included items whose responses
differentiated psychiatrically diagnosed patient groups from "normal" hospital
visitors. While these initial MMPI scales were targeted on the nosological typol-
ogy originally developed by Klaepel, later MMPI scales became more catholic;
the 550 items in the MMPI proved to be such an extraordinarily fecund source
for scale development that within fifteen years after the publication of the
inventory, more than two hundred MMPI scales were already available (Dahlstrom
& Welsh, 1960), and the next ten years saw the birth of hundreds more.

However, while the enormous popularity of the MMPI led such inventory
developers as Berdie and Layton (Minnesota Counseling Inventory), Gough (Cal-
ifornia Psychological Inventory), and Heist and Williams (Omnibus Per-
sonality Inventory) to borrow generously from the MMPI item pool, only
two of the many post-MMPI inventories (Gough's CPI and Frick's OAIS)
borrowed its external scale construction strategy. Instead, most of the
adjustment inventories published in the 1940s and 1950s were similar to
inventories designed in the 1930s.

Adams and Lepley's Personal Audit, published in 1941, fractionated ad-
justment into the traits of Seriousness-Impulsiveness, Firmness-Indecision,
Tranquillity-Irritability, Frankness-Evasion, Stability-Instability, Tolerance-
Intolerance, Steadiness-Emotionality, Persistence-Fluctuation, and Content-
ment-Worry (see Adams, 1941). Another inventory published in the same
year, Johnson's Temperament Analysis, reflected a similarly bipolar view-
point towards aspects of psychopathology, with scales labeled Nervous-
Composed, Depressive-Gayhearted, Active-Quiet, Cordial-Cold, Sympathetic-
Hard-boiled, Subjective-Objective, Aggressive-Submissive, Critical-Appli-
ciable, and Self-Mastery-Impulsive.

The mass of adjustment scales on the market by the late 1930s led some
psychologists to consider merging old inventories to form a new composite,
a procedure pioneered by Bernreuter in 1932 and given some impetus by the
enormous commercial success of his Personality Inventory. Darley and McNi-
ara (1940) factor-analyzed the scores from Bell's (1935) Adjustment Inven-
tory, Rundquist and Sleito's (1936) Minnesota Scale for the Survey of Opin-
ions, and Williamson and Darley's (1937a, b) Minnesota Inventory of Social
Attitudes (an introversion-extroversion inventory) and extracted five factors;
their resulting five new scales (Morale, Social Adjustment, Family Relations,
Emotionality, and Economic Conservatism) were published in 1941 as the
Minnesota Personality Scale (MPS). As sauce for the gander, Berdie and Lay-
ton later combined three scales from the MPS (which they relabeled Family
Relationships, Social Relationships, and Emotional Stability) with four scales
from the MMPI (which they called Conformity [Pd], Adjustment to Reality
[Saf], Mood [D], and Leadership [S8]), and in 1952 published the combined
item pool (the language adapted slightly for high school students) as the Min-
nesota Counseling Inventory.

Inventory "merging" reached its zenith, however, in the continued efforts
of a group of Berkeley psychologists to construct the Omnibus Personality
Inventory (OPI). Their conglomerate had its origins in the Vassar College At-
titude Inventory (VCAI), assembled early in the 1950s from items developed
by Sanford, Barron, and Gough at Berkeley's Institute of Personality Assess-
ment and Research (IPAR) and then revised at Vassar by Sanford, Webster,
and Freedman. Darley and McConnell later merged items from the VCAI
with those from other inventories, and Heist and Williams prepared Forms A
and B of the OPI in 1957; 733 items borrowed from at least four inventories
(MMPI, CPI, Minnesota T-S-E-Inventory, and the VCAI) yielded scores on 18
scales. Form O, a popular 1959 abridgment, yielded scores on five scales
(Thinking Introversion, Theoretical Orientation, Conformity, Originality, and
Estheticism). Forms C and D, two 1963 revisions, included 16 and 12 scalesespectively. Form F, published by Heist and Yonge in 1958, contained 385
items and 14 scales (Thinking Introversion, Theoretical Orientation, Esthetic-
ism, Complexity, Autonomy, Religious Orientation, Social Extroversion,
Impulse Expression, Personal Integration, Anxiety Level, Altruism, Practical
Outlook, Masculinity-Femininity, and Response Bias). Interestingly, after all
these revisions, approximately 25 percent of the 385 OPI items are essentially
the same as items from the MMPI or CPI.

As early as the mid-1930s, when Flanagan (1935) was demonstrating that
the four Bernreuter scales could be factored by two, Guilford began a series
of factor-analytic investigations of personality scales. His 1934 Nebraska
Personality Inventory included one adjustment scale (Emotionality), plus a So-
cial Introversion and a Masculinity scale. Factor analyses of previous intro-
version-extroversion scales led to Guilford and Guilford's Inventory of Factors
STDCR in 1940, with scales (see Guilford & Guilford, 1934, 1936, 1939a, b)
called Social Introversion, Thinking Introversion, Depression, Cycloid (mood
fluctuation), and Rhythmicity (carefree liveliness and impulsivity). By 1943
Guilford and Martin had constructed a number of factorially-based person-
ality measures, which they published as the Inventory of Factors GAMIN
and the Personnel Inventory. The former included two more adjustment scales
(Inferiority Feelings and Nervous Tenseness), along with scales to measure
Pressure for Overt Activity, Social Ascendancy, and Masculinity (Martin,
1945). The latter, aimed at detecting the potential "troublemaker" in indus-
trial settings, included scales for Objectivity, Agreeableness, and Cooperation.
A condensation and revision of all three Guilford inventories, Guilford and
Zimmerman's Temperament Survey, was published in 1949.

During the same period, Cattell was undertaking a factorial investigation of
the "total personality sphere," an endeavor which culminated in the 1949 pub-
lication of the Sixteen Personality Factor Questionnaire (16PF). Of the six
scales in the 16PF which had been shown to differentiate between neurotic and
normal groups, three had high loadings on the broad second-order "Anxiety" factor
of the 16PF (see Cattell, Eber, & Tatsuoka, 1970). Scheier and Cattell com-
bined these three scales into one Anxiety score, which was supplemented by
the other three differentiating scales (Tender-Mindedness, Depression, and Sub-
missiveness) and published in 1961 as the Neuroticism Scale Questionnaire.
A second inventory by Cattell and Scheier, the IPAT Anxiety Scale Questionnaire,
measured only components of the second-order anxiety factor and provided
five scales (Self-Sentiment Development, Ego Strength, Paranoid Trend, Guilt
Proneness, and Ergic Tension), as well as scores for Covert and Overt anxiety, plus a total anxiety score.

While Guilford and Cattell had been developing factor scales on normal subjects and then testing whether the resulting scales differentiated psychiatric patients from other groups, Eysenck had been constructing factor scales aimed at doing this job from the start. Eysenck and his associates constructed two broad factor scales (Neuroticism and Psychoticism), only the first of which has been incorporated into published inventories. The 1959 Maudsley Personality Inventory (MPI) provided scores for Neuroticism and Extroversion-Introversion. While Bernreuter’s Neurotic Tendencies and Introversion-Extroversion scales had correlated .96, the two analogous MPI scales correlated no higher than .30. However, even this modest degree of relationship stimulated Eysenck to revise the scales; the resulting Eysenck Personality Inventory, which included the same two scales (now with virtually no correlation between them), was published in 1964.

While the factor-analytic triumvirate (Cattell, Eysenck, and Guilford) attempted to map personality structure more broadly, a number of other inventory developers were using variants of the same internal strategy to construct specific adjustment scales. For example, Haston had used internal consistency analysis (though not factor analysis) to develop his Personal Adjustment Inventory. This inventory, published in 1949, provided homogeneous and reasonably independent scales aimed at the measurement of Emotional Stability, Confidence, Personal Relations, Home Satisfaction, and Analytical Thinking. Around the same period, Gordon utilized factor analysis and a forced-choice item format to isolate four scales (Emotional Stability, Responsibility, Ascendancy, and Sociability), which he published in 1953 as the Personal Profile. A parallel inventory of Gordon’s, the Personal Inventory, was published in 1956 and provided scales for Personal Relations, Vigor, Cautiousness, and Original Thinking.

Although the internal strategy of scale construction has dominated other approaches of late, two important post-World War II inventories were developed from the external strategy. The most popular of these was Gough’s 1956 California Psychological Inventory (CPI), which borrowed about two-fifths of its items from the MMPI. Eleven of the 18 original CPI scales included items which differentiated between groups of subjects who had been rated (or could be logically seen) as being at extreme poles of the particular trait targeted for the scale. While the CPI was explicitly designed to measure important individual differences not tapped by the host of adjustment inventories on the market at that time, a few CPI scales (e.g., Socialization, Self-Control, Self-Acceptance, Flexibility) appear to tap traits within the adjustment domain.

Unfortunately, inventories constructed by the external strategy have tended to provide highly correlated scales; a host of correlational and factor-analytic studies of SVIB, MMPI, and CPI scale scores attest to the relative redundancy within each of these three inventories, a redundancy which is aggravated by the fact that the same items are scored on more than one scale. For example, the SVIB Chemist and Physicist scales correlated .93 (Strong, 1959), the MMPI Hypochondriasis and Hysteria scales correlated .98 among normal subjects (Thumim, 1969) and .81 among psychiatric patients (Goldberg, 1965), and the CPI Dominance and Sociability scales correlated .97 among females and .61 among males (Gough, 1957). Though most of these correlations are close to the maximum possible, given the reliabilities of the scales, only the first approaches that pinnacle reached by two of the four original Bernreuter scales.

The one exception to this general rule was Fricke’s Opinion, Attitude, and Interest Survey (OAIS), the manual for which was released in 1963. Fricke included in each of the OAIS scales only those items whose validity had been established against non-test criteria, and whose correlations with other items were such that the resulting scales would be minimally intercorrelated. Only two of the OAIS scales (Social Adjustment and Emotional Adjustment) fall in the adjustment domain, and their intercorrelations within various samples ranged from about .10 to .40 (Fricke, 1963); the other OAIS scales, which measure aspects of vocational interests and scholastic potential, will be discussed later.

The use of some mixed strategy of inventory construction, exemplified by Fricke’s use of both external and internal criteria for item inclusion, is not new. While the very first adjustment scales were constructed using a purely intuitive approach, most scale developers began using a mixture of intuitive and internal strategies as early as the 1920s. A typical procedure was to assemble an item pool on purely intuitive grounds, administer the pool to a sample of subjects, and then correlate responses to each item with total (a priori) scale scores; items not significantly associated with total scores were then eliminated from the refined measure. Such a procedure is directly analogous to that of selecting items with high loadings on the first unrotated factor when responses to a set of items purportedly tapping some single dimension are subjected to factor analysis.

The epitome of this mixed intuitive-internal strategy of inventory construction can be found in three recent inventories: Edwards’ (1966) Personality Inventory, Jackson’s (1967) Personality Research Form, and Jackson and Messick’s as yet unpublished Differential Personality Inventory (DPI). Of these three new multi-scale inventories, the first two are not oriented toward the adjustment domain, and therefore only the DPI will be discussed at this point. While items were initially assembled for each of the DPI scales on intuitive grounds, only those items which had “high content saturation” on their respective scales and low correlations with “stylistic” and other content scales were retained (internal strategy). Moreover, all scales were composed of equal numbers of true and false keyed items. The resulting inventory includes scales for Cynicism, Depression, Familial Discord, Health Concern, Hostility, Impulsivity, Irritability, Neurotic Disorganization, Psychotic Tendencies, Religiousness, Socially Deviant Attitudes, Somatic Complaints, Defensiveness, Insomnia, Broodiness, Desocialization, Thinking Disorganization, Feelings of Unreality, Hypochondriasis, Ideas of Persuasion, Mood Fluctuation, Panic Reaction, Perceptual Distortion, Repression, Sadism, Self-Depreciation,
XIV. PERSONALITY SCALES AND INVENTORIES

Technology in Pittsburgh. Approximately one thousand interest items were developed in a seminar of Yoakum's in 1919, and subsets of these items were used by many investigators over the next fifty years. In 1921 Moore reported a study of the differential interests of sales vs. design engineers, and one year later Freyd developed some preliminary interest scales for engineers. The same year Miner, who apparently originated the idea of comparing the responses of individuals in a specific occupational group with a sample from diverse occupations (Campbell, 1968), published some interest scales based upon the responses of students tested in Pittsburgh high schools beginning in 1918.

By the early 1920s, a number of the younger members of the Carnegie group had been exploring the differential interests of individuals in various occupations. Ream investigated successful vs. unsuccessful life insurance salesmen, and Cowdery compared the interests of physicians, lawyers, and engineers. By far the most illustrious of this group, however, was E. K. Strong, who, after moving from Carnegie to Stanford in 1923, began to develop his Vocational Interest Blank (SVIB), which was published in 1927. The story of Strong's involvement in interest measurement and of the resulting SVIB (including all its forms and revisions) can be found in Campbell (1968). The approach to interest inventory construction popularized by Strong included at least four salient features: (a) the use of occupational, educational, and avocational preference items from the original Yoakum pool, most of these presented as a single stimulus; (b) the use of a "Like-Indifferent-Dislike" (L-I-D) response option with most of the items; (c) the use of an external strategy of scale construction (keying only those items which had been shown empirically to differentiate among occupational groups); and (d) scale development based upon the comparison of a particular occupational group with a sample of "men in general." While all of these four features of the SVIB influenced later interest inventories, most interest developers dropped one or more features in designing their own products.

One of the early alternatives to the SVIB was Garretson and Symonds' Interest Questionnaire for High School Students; this 234-item inventory (with a L-I-D response format) was first published in 1930 and revised in 1942. In 1932 Brainard and Stewart published their Specific Interest Inventory; 100 work activities were included, each with a 5-point rating scale (from "like it very much" to "dislike it very much"), intuitively grouped to provide 20 "mode of activity" scores (e.g., Physical Work, Mechanical Work, Vocal Expression, Experimenting, Creative Imagination, etc.). The inventory was revised by Brainard and Brainard in 1945 and renamed the Occupational Preference Inventory; 140 items (with a 5-point response format) yielded intuitively constructed scores for 28 occupations, grouped into seven families. The inventory was revised once again in 1956 to yield scores on Commerical, Mechanical, Professional, Esthetic, Scientific, Personal Service (girls), and Agriculture (boys) interest scales.

By far the most important competitor to the SVIB, however, was Kuder's Preference Record—Vocational, published in 1934. Providing measures of Scientific, Computational, Musical, Artistic, Literary, Social Service,
Persuasive interests, this inventory originally included 330 items, each a paired comparison between two occupational or avocational activities. Items were clustered on the basis of homogeneity analyses, and the resulting seven scales were reasonably independent of each other. In later versions of this inventory, Kuder switched the format to item triads, with instructions to rank order the three items within each triad. A 1946 revision of the inventory added a Mechanical and a Clerical interest scale, plus a measure of Masculinity-Femininity. A 1956 revision added an Outdoor interest scale. In 1964, a parallel inventory for a slightly younger audience was published as Kuder’s General Interest Survey.

While Kuder utilized a popular variant of the internal strategy of scale construction, he did not use factor analysis. It remained for Thurstone to apply factor-analytic procedures to interest items; his Vocational Interest Schedule (VIS) was developed during the same period as Kuder’s inventory and published in 1935. Thurstone’s VIS, which contained as items 72 occupational titles, each presented with a L-I-D response format, provided scales of Commercial, Academic, Scientific, Biological, Legal, Athletic, and Descriptive (humanistic) interests. By 1947, when a revised version of the inventory was published as the Thurstone Interest Schedule, Thurstone himself had shunned factor analysis in favor of a modified internal consistency procedure. The 1947 Thurstone inventory included as items 100 pairs of occupational titles, with instructions to indicate one’s preference within each forced-choice pair. Ten scores were provided: Physical Science, Biological Science, Computational, Business, Executive, Persuasive, Linguistic, Humanitarian, Artistic, and Musical.

A somewhat similar interest inventory was published by Le Suer in 1937; his Occupational Interest Blank included 100 occupational titles as items with a L-I-D response format. In contrast to the VIS, however, Le Suer grouped his items on a purely intuitive basis to construct Professional, Technical, Clerical, Sales, Artistic, Skilled Trades, Semiskilled Trades, and Adventurous scales.

The original set of interest items spawned by Yoakum’s 1919 seminar was used during the early 1930s to develop Cleeton’s Vocational Interest Inventory, which was published in 1937. Cleeton’s inventory included 630 of these items, presented in a “like-dislike” response format, plus 40 items designed to measure Social Adjustment. The men’s form of Cleeton’s inventory provided scores for Biological Science, Physical Science, Social Service, Sales, Business Administration, Financial, Legal-Literary, Mechanical, and Creative occupations; the women’s form for Natural Science, Social Service, Sales, Office Work, Personal Service, Mechanical, Creative, Teaching, and Household vocations. A 1943 revision of these two inventories added an Agricultural scale to the men’s form, and replaced the Teaching and Household scales on the women’s form with scales for Grade School Teaching, High School and College Teaching, Housekeeper-Factory Worker, and Homemaking-Child Care. The Social Adjustment scale was dropped from both forms.

Of all the departures from the Yoakum item pool, the most interesting was that of Cardall, who engaged 106 businessmen to keep work diaries reflecting their hour-to-hour activities. From these diaries, Cardall extracted over two thousand descriptions of business activities (e.g., “pay bills and bring back receipts”), which he then reduced on intuitive grounds to a smaller subset. A cluster analysis of these descriptors formed the basis for Cardall’s five-scale Primary Business Interests Test (PBIT), which was published in 1942; the PBIT provided interest scores for Accounting, Collections and Adjustments, Sales-Office, Sales-Store, and Stenographic-Filing activities.

A potpourri of over 400 information, preference, biographical, and completion items were included in Gentry’s Vocational Inventory, which was published in 1940; the items were grouped into scales for Social Service, Business, Law-Government, Art, Mechanical Designing, Mechanical Construction, Science, and Literary interests—plus an Introversion-Extroversion scale. Moreover, a spate of similar interest inventories was also published during the 1940s, a symptom of the seemingly voracious market for putative measures of vocational satisfaction and success. Most of these are of interest today only insofar as they demonstrate the sort of items utilized during this period and the kinds of individual differences deemed important for forecasting vocational success. One of these, Kobel et al.’s Inventory of Vocational Interests, was published in 1943 and furnished Mechanical, Academic, Artistic, Business, and Agricultural interest scores. The inventory included 25 questions (e.g., “What would you like to do best with an airplane?”), each presented with 10 alternatives (e.g., “work on it as a mechanic, pilot it, study its history, write a poem about it, advertise its uses,” etc.), with instructions to check 3 of the 10 alternatives. A revision of this inventory, published in the late 1950s, used the same sort of items and furnished the same five scores.

Another such attempt was Lee and Thorpe’s Occupational Interest Inventory, also published in 1943. This inventory included 240 paired-comparisons activity items, 40 in each of six fields; the items within each field were further stratified by three levels of skill and responsibility, and by three broad types of interest modes. The resulting scales generated scores for six interest fields (Social, Natural, Mechanical, Business, Arts, and Sciences) and three interest types (Verbal, Manipulative, and Computational), plus an overall Interest Level. A 1956 revision of this inventory retained these same features. Another type of item stratification was used in Van Allyn’s Job Qualification Inventory, first published in 1945. Within each of 35 occupational areas, one item dealt with preferences, another with past performance, a third with education, a fourth with vocational aspiration, a fifth with paid experience, and a sixth with indications of unusual proficiency; the resulting 210 items were presented in a NO?-YES response format and yielded 35 occupational scores.

A revision of this inventory in 1958 was titled the Qualifications Record and yielded 45 occupational scores, grouped into seven broad families. A shorter version of the Qualifications Record with the same 45 scales was published as the Career Finder in 1960.

In contrast to these intuitive scale construction procedures, at least one inventory developer in the 1940s returned to the external strategy pioneered by Strong. Beginning in the late 1930s, Gregory administered about 900 interest items to a number of samples of college students; the responses of
juniors and seniors majoring in each of 28 college curriculum areas were contrasted with the responses of a sample of college freshmen, and the most differentiating items were retained to form Gregory's Academic Interest Inventory, which was published in 1946. The 300-item inventory utilized a five-category response format and yielded 28 academic interest scores (Agriculture, Architecture, Biology, ... Religion, Secondary Education, Sociology, Speech). A somewhat similar inventory, Baldwin's Motivation Indicator, was constructed by the intuitive strategy and aimed at the high school rather than the college curriculum. This inventory, published in 1947, included 81 curricular activity statements (nine in each of nine areas) and 25 social activity statements (five in each of five areas). All items were presented in sets of four, with instructions to select the two most liked; each item was presented four times (and compared with eight other items). The inventory furnished nine curricular scores (Biological Sciences, Physical Sciences, Social Sciences, Literary Arts, Graphic Arts, Industrial Arts, Agricultural Arts, Clerical-Verbal, and Clerical-Numerical) and five motivational scores (Altruistic, Promotional, Administrative, Distributive [sales], and Creative).

In a more recent attempt to predict college curricular goals, Fricke (1963) used a mixture of the external and internal strategies of scale construction to develop five interest scales (Business, Humanities, Social Science, Physical Science, and Biological Science) for his Opinion, Attitude, and Interest Survey (OIAS). In marked contrast to Gregory, however, Fricke elected to eschew college major as a criterion in favor of a one-item ranking of the five curricular areas. Each of the five OIAS interest scales included about 100 items, each of which differentiated between students who ranked a particular area first or second versus those who ranked it fourth or fifth, when asked to rank order their preferences for the five fields. Moreover, although Fricke simultaneously attempted to keep his scales as independent as possible, about half of the correlations among the five OIAS interest scales were above .30.

By the end of World War II, Guilford and his associates had moved into interest measurement. A pool of 540 activity items (with a "Dislike—Like as Hobby—Like as Vocation" response format) was reduced to 360 items by internal consistency analyses and published in 1948 as the Guilford-Schniedman-Zimmerman Interest Survey. Each of nine interest areas was subdivided into two components: Artistic (Appreciative and Expressive), Linguistic (Appreciative and Expressive), Scientific (Investigatory and Theoretical), Mechanical (Manipulative and Designing), Outdoor (Natural and Athletic), Business-Political (Mercantile and Leadership), Social (Persuasive and Gregarious), Personal Assistance (Personal Service and Social Welfare), and Office Work (Clerical and Numerical); each of the resulting 18 scales, in turn, could be scored both for Hobby and for Vocation, for a total of 36 scores in all. The Interest Survey should not be confused with the Guilford-Zimmerman Interest Inventory, a 1962 publication of Joan Guilford and the same Zimmerman. The latter inventory included 150 items (with a four-category response format), 15 for each of 10 scales (Mechanical, Natural, Aesthetic, Service, Clerical, Mercantile, Leadership, Literary, Scientific, and Creative).

A survey of factor-analytic studies of interest and personality inventories convinced Kuder of the need to measure seven additional dimensions beyond those already included in his Preference Record—Vocational (PR—V), and scales for five of these dimensions were included in the numerous revisions of his Preference Record—Personal (PR—P) during the 15 years after its publication in 1948. The scales in the first edition were labeled Sociable, Practical, Theoretical, Agreeable, and Dominant, though these were soon changed to preferences for Group Activity, Stable Situations, Working with Ideas, Avoiding Conflict, and Directing Others. While Kuder used a variant of the internal strategy to construct his PR—V and his PR—P, by the middle 1950s the enormous popularity of the SVIB stimulated him to develop another competing interest inventory using the external strategy of scale construction. Kuder's Preference Record—Occupational (PR—O), which was published in 1956, contained 100 sets of item triads with instructions to rank-order the items in each triad; items whose responses differentiated between a particular occupational group and a sample of men from many occupations were keyed to provide 22 occupational scores. Like the SVIB, the number of occupational scales available from the PR—O has been gradually increasing over the years; by the time the inventory was revised in 1963, 51 occupational scales were available.

In contrast to the careful scale construction procedures employed in the inventories developed by Kuder, a number of other interest inventories published in the 1950s were practically indistinguishable from some much earlier models. For example, Jenkins' How Well Do You Know Your Interests?, which was published in 1957, was at least faintly reminiscent of Travis' Diagnostic Character Test, which was published in 1925. While the former included 54 scales and the latter 50, 46 of the scales in Jenkins' inventory (and all in Travis') were composed of only two items. However, Jenkins used an internal strategy of scale construction, and he selected the two items with the highest factor loadings from a much larger pool; 7 of the 54 interest scales included four items, and 1 longer scale was provided as a measure of masculinity-femininity. All items were presented with a 6-category response format (from "like tremendously" to "dislike tremendously"). Another such inventory from the same period was Curtis' Interest Scale, published in 1959. This throwback employed 5 items, each with 10 alternatives (one from each of 10 interest areas), with instructions to rank-order the 10 alternatives; the inventory provided scales for Business, Mechanics, Applied Arts, Direct Sales, Production, Science, Entertainment, Interpersonal, Computation, and Farming, plus a five-item scale to measure Desire for Responsibility.

The differences among item types included in all these inventories may have served to mask their one common characteristic—namely, their reliance on verbal statements or verbal titles. Consequently, all of these inventories could only be used with individuals who were neither culturally nor educationally disadvantaged. However, as society began to recognize the need to help its more handicapped members, the demand for nonverbal items began to increase. In response to these pressures, two Picture Interest Inventories (PII) were published, the first by Weingarten in 1958 and the second by Geist in 1959. Weingarten's PII was modeled after Lee and Thorpe's Occupational
Interest Inventory and furnished scales relabeled Interpersonal Service, Natural, Mechanical, Business, Esthetic, Scientific, Verbal, and Computational, as well as a scale to measure Time Perspective. Geist's PII was modeled after Kuder's Preference Record—Vocational and included 11 interest scales (Persuasive, Clerical, Mechanical, Dramatic, Musical, Scientific, Outdoor, Literary, Computational, Artistic, and Social Service) and six motivational scales (Family, Prestige, Financial, Intrinsic, Environmental, and Past Experience). The men's form of Geist's inventory included 132 line drawings of various activities arranged in 44 triads; instructions were to choose the most liked drawing in each triad. A 1964 revision of Geist's PII added a Personal Service scale to the women's form, which included 27 pictorial triads.

Societal pressures to provide counseling services for handicapped and deprived individuals also led to the development of a few recent interest inventories aimed at lower-level positions in the work hierarchy. In contrast to the great mass of measures targeted for the college-oriented, Gordon's Occupational Check List was aimed directly at individuals with high school educations or less. This 1961 inventory, modeled after Mooney's Problem Check List, consisted of 240 task descriptions with instructions to undertake the tasks one liked and to circle the tasks one liked best. The items were keyed via internal consistency procedures to provide five scales: Business, Outdoor, Arts, Technology, and Service. A second instrument directed at the same audience, Clark's 1965 Minnesota Vocational Interest Inventory, utilized the external strategy of scale construction to provide occupational scales for lower-level jobs than those included in the SVIB.

SCHOLASTIC PREDICTORS

The preceding attempts to predict occupational satisfaction and attainment are closely related to the development of measures aimed at the prediction of success in various vocational training programs and in colleges and universities. As already noted, one of the earliest applications of the external strategy of scale construction was Pressey's (1921) attempt to differentiate students who earned satisfactory grades in college from their less successful peers by means of their responses to items from his X-O tests. However, most research on scholastic prediction has relied heavily upon aptitude measures, and few early attempts to develop personality scales for use in this area can be found. Moreover, the very success of psychologists' attempts to predict grade point average (GPA) by means of scholastic aptitude tests appears to have led investigators to focus research on measures of more specific aptitudes rather than to pursue the search for nonintellective sorts of predictors.

One of the early exceptions to this trend can be found in the work of Wrenn, who administered 69 items relating to study habits, each with a three-category response format (rarely—sometimes—often), to 220 Stanford students, half of whom were in the top 10 percent and half in the lowest 20 percent of the distribution of college grades; the students in the two groups were matched on IQ test scores. The 30 items which significantly differentiated between these groups formed Wrenn's Study Habits Inventory, which was published in 1934. A 28-item revision was published in 1941.

Dunlap used the same external strategy and 435 interest items to construct scales to predict the subscores of the New Stanford Achievement Test, the Metropolitan Achievement Tests, and the Terman Group Test of Mental Ability; the 100 items with the most significant correlations with these test scores were incorporated into Dunlap's Academic Preference Blank, which was published in 1937. The items were presented with a "Like-Indifferent-Dislike-Unknown" response format, and scores could be obtained for History, Geography, Arithmetic, Literature, Language Usage, Paragraph Meaning, and Word Meaning subscores, plus General Achievement, Mental Age, and IQ scores. Later editions of this unique inventory reduced the item pool to 90 items to facilitate machine scoring.

In the 1940s, the major thrust of scholastic prediction research centered on the construction of differential aptitude tests, a development stimulated by the apparent success of aptitude batteries in selecting military personnel in World War II. While a number of investigators began to assess the incremental validity of existing adjustment, interest, and introversion- extraversion scales in predicting college achievement, there were relatively few attempts to develop new personality measures for this specific purpose. Traxler published an 85-item Survey of Study Habits in 1944, which was intended primarily as a means of student self-analysis in academic counseling settings. However, the major development during this decade was the publication in 1949 of Bow's College Inventory of Academic Adjustment (CIAA), which provided six subscores (Curricular Adjustment, Level of Aspiration, Use of Time, Study Skills, Mental Health, and Personal Relations) plus a total score. The 90 items in this inventory, presented with a "Yes-No" response format, included only those items (from a much larger set initially administered) which differentiated academic overachievers from underachievers at Pennsylvania State College; the differentiating items were grouped intuitively into the six subscales, whose median intercorrelation was about .45. Bow's method of inventory construction insured that CIAA scores had quite low correlations with those from aptitude tests, yet correlated between .30 and .40 with college GPA. Consequently, an admissions officer using this inventory along with an aptitude test battery might be able to raise the correlation with GPA from around .50 (tests alone) to around .60 (tests plus CIAA), the latter value being approximately equal to the predictive validity of high school grades.

The 1950s saw an emerging interest in forecasting academic success via structured personality measures. While critics of aptitude testing had long maintained that certain students obtained lower aptitude scores because of intense "test anxiety," it was not until the 1950s that any systematic research was begun to measure this putative personality trait. One of the early attempts was made by Mandler and S. B. Sarason (1952) at Yale University, culminating in their Test Anxiety Questionnaire. A psychometrically more polished instrument, titled the Test Anxiety Scale, was fashioned by I. G. Sarason (1958) at the University of Washington. While both of these measures provided single scores for test anxiety, Alpert (now Baba Ram Dass) and Haber (1960)
fractionated the construct into two components—Facilitating Anxiety and De-
bilitating Anxiety (postulated to affect examination performance in opposite
directions)—in their Achievement Anxiety Test, developed at Stanford Uni-
versity. Interestingly, no one has yet demonstrated unambiguously that any
of these anxiety scales function as significant suppressor or moderator vari-
ables in the prediction of college grades from aptitude tests, or that their
inclusion in a college admissions battery significantly improves the level of pre-
dictive validity beyond that achievable by high school grades and aptitude
tests alone.

The test anxiety scales, which were constructed by the intuitive assembly
of items (sometimes refined by internal consistency procedures), furnish the
most salient exceptions to the use of the external strategy in constructing
academic prediction scales. The early scales developed by Friesey, Wrenn,
Dunlap, and Bowe were based upon external criteria, as were the later ones
developed by Brown and Holtzman, Gough, and Fricke. Brown and Holtz-
man’s Survey of Study Habits and Attitudes (SSHA), published in 1953, in-
cluded 75 items referring to various study attitudes, culled from a larger set
of statements transcribed from student interviews; separate (but overlapping)
males and females SSHA prediction scales included those items which differen-
tiated college overachievers from underachievers. Both scales had quite low
relationships with scholastic aptitude test scores and correlated about .45 with
college grades. The SSHA was revised in 1966 and 25 items were added; the
100 items were intuitively grouped into four highly correlated 25-item scales
(Delay Avoidance, Work Methods, Teacher Approval, and Educational Ac-
ceptance), the first two of which could be combined to form a scale of Study
Habits and the last two a scale of Study Attitudes—both of which, in turn,
could be combined to furnish a total score, dubbed Study Orientation.

A somewhat similar fractionation of study attitudes can be found in Car-
ter’s California Study Methods Survey (CSMS), which included three highly
correlated subscales (School Attitudes, Study Mechanics, and Planning) plus
a total score. This 1958 inventory consisted of 150 questions, presented with a
“Yes-No” response format. Like other scholastic inventories, its total score
correlated about .45 with college grades. The CSMS should not be confused
with Gough’s California Psychological Inventory (CPI), which was published
in 1956. Three CPI scales were specifically constructed as measures of achieve-
ment potential and intellectual efficiency: Achievement via Conformance
(Ac), composed of items significantly correlated with high school grades;
Achievement via Independence (Ai), which included those items significantly
correlated with college grades; and Intellectual Efficiency (Ie), composed of
items significantly correlated with intelligence test scores. Over the years,
the validity of these CPI scales has been assessed, both singly and in combina-
tion, in a host of settings ranging from high schools (e.g., Gough, 1964, 1966a)
to medical schools (e.g., Gough, 1967; Gough & Hall, 1964).

As already noted, diverse scales constructed by the external strategy often
display rather high intercorrelations; in the case of the CPI, the Ac and Ai
scales intercorrelated about .40, and they both correlated around .55 with
Ie (Gough, 1957). Fricke (1963) attempted to construct three scholastic

predictors for his Opinion, Attitude, and Interest Survey (OAI) which would
be independent of each other (and of the other scales in the OAI). The OAI
Achiever Personality scale (Ach P) included 86 items whose responses differen-
tiated college overachievers from underachievers; the Intellectual Quality
scale (Int Q) included 85 items with significant correlations with aptitude test
scores; and the Creative Personality scale (Cre P) included 101 items differen-
tiating students nominated as being unusually creative by their college
instructors from students nominated as being low in creativity. The intercor-
correlations among these three OAI scales were indeed quite low, as were
their correlations with the three CPI scholastic scales; while Int Q correlated
about .55 with Ai and about .35 with Ie, the other cross-inventory correla-
tions ranged from zero to .30 (Fricke, 1963). While each inventory developer
has interpreted this correlational pattern as evidence for the relative utility of
his own scales, no independent investigator has yet tested the comparative
validity of the CPI and OAI in forecasting academic promise.

All of the scholastic scales and inventories discussed so far were constructed
as general predictors of academic promise, each developed in the expectation
that it would provide reasonably uniform predictions across various types of
courses and diverse curricular areas. In contrast, Siegel and Siegel (1965) have
developed a personality scale specifically as a differential predictor of schol-
astic promise. Their Educational Set Scale (ESS) was composed of 31 item
triads (e.g., “Assume you are enrolled in an English course and must learn
about the following: (a) the dates and major works of well-known poets; (b)
the role of the playwright in contemporary society; (c) the structure of sonnets”),
with instructions to select the most preferred and the least pre-
ferred alternative. Within each triad, one statement described a task requiring
the acquisition of factual information, and another described a more con-
ceptual learning task; the resulting ESS score was designed to measure students’
preferences for conceptual versus factual knowledge and thus to relate differen-
tially to academic performance under differing instructional procedures.
Unfortunately, the evidence for the differential validity of the ESS is far from
compelling (e.g., Goldberg, 1969).

By far the most ambitious attempt to develop scholastic prediction scales
may be found in the unpublished Academic Behavior Inventory (ABI), con-
structed by Giddan, King, and Lovell. Over 850 items (at least a quarter of
which were borrowed from the MMPI, CPI, SVIB, OPI, EPPS, ACL, OAI,
PRF, and other inventories) were included in various forms of the ABI and
administered during the 1960s to diverse student groups. Form E of the ABI
included 458 items, presented in a true-false response format. Fifteen ABI
scales were constructed by internal consistency analyses and grouped under
five rubrics: Subjective Distress (Manifest Anxiety, Alienation, and Test
Anxiety); Extrinsic Motivation (Achievement, Persistence, and Extrinsic Reward);
Intrinsic Motivation (Affiliation, Class Participation, and Intrinsic Reward);
Creativity (Originality, Flexibility, and Orderliness); and Ideology (Liberalism-
Conservatism, Social Activism, and Sexual Permissiveness); these 15 constructs
were selected to include the most salient traits purportedly tapped by past
scholastic prediction scales as well as conceptions characterizing the more
recent concerns and complaints of college students. Nine other ABI scales were constructed by the external strategy, including a Masculinity-Femininity measure, plus separate male and female scales developed to predict Academic Performance, Overachievement, Academic Capacity, and Academic Motivation.

INTROVERSION-EXTROVERSION

As already noted, the diverse measures of psychopathology, vocational interests, and scholastic potential which have proliferated over the years can be viewed as responses to societal pressures upon psychologists to forecast significant personal outcomes. The measures included in the remaining columns of Table 1, on the other hand, are not so easily viewed as reactions to applied demands. Moreover, most of these measures appear to have been spawned by a few highly influential conceptions of the structure of individual differences.

Perhaps the most compelling of all such theoretical constructs has been that of introversion-extroversion (I-E), popularized in the theoretical writings of the psychoanalytic rebel Carl Jung about the time of World War I and extended in the writings of the psychologist William McDougall in the early 1920s. From this period on, probably more effort has been expended in attempts to construct measures of I-E and the related construct dominance-submissiveness than any trait complex other than intelligence and adjustment. Conklin (1923) published an early review of the I-E literature, and the next year Freyd (1924) published a list of 54 types of behaviors purportedly related to the I-E construct. This list formed the basis for the development of many early I-E scales, including those by Heidbreder (1926) and Root (1931).

One of the first I-E scales was Laird's Colgate Mental Hygiene Inventory—Form C, which was published in 1925; a shortened version was reported by Whitman (1929). Laird's primary interest lay in developing a reliable index of adjustment, and he conceived of the I-E construct as an important determinant of behavioral differences among psychotic patients. This viewpoint, which enjoyed wide popularity in the 1920s, influenced the development of Travis' (1925) Diagnostic Character Test. In an early unpublished study, Bathurst showed that schizophrenics could be differentiated from manic-depressive patients on I-E scales, the former achieving more introverted and the latter more extroverted scores. In a 1927 master's thesis, Kohlstedt administered 100 putative I-E items to 100 schizophrenic and 100 manic-depressive patients; the 50 most highly differentiating items were included in Neyman and Kohlstedt's (1929) New Diagnostic Test for I-E. Gilliland and Morgan (1931) then administered these 50 items to new samples of 60 schizophrenic and 65 manic-depressive patients; the 35 most differentiating items were included in their Northwestern University I-E Test. As evidence for the validity of their scale, Gilliland and Morgan (1931) reported a bimodal distribution of I-E scores among psychiatric patients (most of the schizophrenics falling within the introverted mode and most of the manic-depressives falling within the extroverted mode) and a normal distribution of scores among college students; in the latter sample, the scale correlated about .50 with peer ratings of I-E.

One of the first scale developers to object to this criterion for the construction of I-E scales was Conklin (1927), who attempted to develop an I-E measure within normal samples. Conklin eschewed the Freyd (1924) list in favor of 100 activity items (e.g., playing baseball, hearing lectures, talking with friends, reading essays, etc.), half judged as likely to be preferred by extroverts and half by introverts. About 350 college students responded to each item on a 9-point scale, and those subjects scoring in the top and bottom 10 percent of the distribution of a priori scale scores were used as two criterion groups for item analysis. The 40 most differentiating items (20 preferred by introverts and 20 by extroverts) were included in Conklin's (1927) E-I Interest Ratio. As one might expect from a scale constructed by this variant of the internal strategy, the reliability (homogeneity) of the E-I Interest Ratio was over .90 in new samples. A similarly constructed measure developed during the same period was Bernreuter's (1933a) Self-Sufficiency scale. Bernreuter administered 132 items, each purportedly tapping individual differences in dependency upon other persons, to 127 college students with a "Yes-No" response format. An a priori key was used to provide two samples with extreme scores, which were used for all item analyses. Sixty items were included in the revised scale, which had reliability (homogeneity) values around .85.

In contrast, Floyd and Gordon Allport used a variant of the external strategy in constructing their Ascendance-Submission Reaction Study, which was first published in 1928 and revised in 1939. The Allports developed a set of items describing hypothetical situations (e.g., "Are you embarrassed if you have greeted a stranger whom you have mistaken for an acquaintance?"); each presented with two or three alternative answers (e.g., "very much, somewhat, not at all"); 41 of these items were administered to 400 college males, and 49 items to 200 college females, all of whom rated themselves and were rated by four friends on an A-S rating scale. Scoring weights for each item alternative were constructed on the basis of the self- and peer-rater criteria. Scale validities against peer ratings in new samples averaged about .30, and reliability (homogeneity) coefficients averaged about .75 (Allport, 1928).

As already noted, 1932 saw the publication of Bernreuter's Personality Inventory, which included scales constructed to predict Laird's I-E scale, the Allports' Ascendance-Submission scale, the Thurstones' Neurotic Tendencies scale, and Bernreuter's own Self-Sufficiency scale. While an armchair analysis of the scale labels might have clustered the first two in one group (I-E) and the second two in another (Adjustment), Flanagan's (1935) factor analysis of the four scores produced a radically different alignment; the I-E measure turned out to be virtually indistinguishable from the Neurotic Tendencies scale (the Adjustment factor), while Self-Sufficiency plus Ascendance-Submission clustered as the I-E factor.

While Flanagan (1935) was factoring the four Bernreuter scales, Guilford and Guilford (1934) were factoring a set of typical I-E items. The Guilfords culled 75 I-E descriptive phrases from the writings of Jung, from Freyd's (1924) list and from the previous I-E scales of Laird, Neyman, and Kohlstedt (1929), and Gilliland and Morgan (1931). They then administered 35
representative items from this set to 930 college students and from the item
correlations tried to ascertain whether there was a general factor of I-E akin
to that found with intelligence test scores. After rejecting this hypothesis, the
Guilford rotated four factors and named them Social I-E, Emotional Sensitivity,
Impulsiveness, and Interest in Self. A 1936 paper reported another factor
analysis of the same items and the same sample; five factors were
rotated and named S (Shyness, Seclusiveness, or Social I-E), E (Emotionality),
M (Masculinity), R (Rhathymia, the Guilford's term for carefree impulsiveness),
and T (Thinking I-E)—the first three of which were incorporated into
Guilford's Nebraska Personality Inventory, which was published in 1934. In
two 1939 papers, the Guilford's reported factorial investigations of additional
items in samples of 1,000 and 600 college students. In one of these studies,
the Guilford's rotated nine factors, among them D (Depression), R (Rhathy-
mia), LT (Liking for Thinking), S (Shyness), and T (Thinking I-E); in the
second study, they rotated seven factors, including N (Nervousness), and GD
(General Drive). These factors were refined (and some were relabeled) in
subsequent factor analyses, and they were included in the Guilford's Inven-
tory of Factors STDCR, which was published in 1940.

In contrast to the factor-analytic procedures utilized by the Guilford's,
Williamson and Darley (1937a, b) used other variants of the internal strategy
of scale construction to develop their Minnesota Inventory of Social Attitudes.
This inventory was composed of two forms, each of which furnished a single
score; the first form consisted of 40 items describing various behaviors and
feelings in social situations (Social Behavior), and the second consisted of 40
items asking for preferences for various types of social relationships (Social
Preferences). The two scores correlated about .45, and each had riability
(homogeneity) coefficients in the .90s.

The Guilford's demonstration that I-E items clustered on at least three to
five factors stimulated Evans and McConnell (1941) into constructing an in-
ventory which would measure three I-E facets—labeled Thinking, Social,
and Emotional I-E—as independently as possible. Their three-scale Minnesota
T-S-E Inventory was published in 1942 and revised in 1957. After having
formulated the characteristics associated with three distinct types of I-E,
Evans and McConnell (1941) wrote 216 items reflecting behavior in each
category and asked ten judges to classify the items; those 197 items which
elicited at least moderate judgmental agreement were supplemented by 43
new ones and administered to about 300 college students. All of the 240 items
were correlated with three a priori scale scores, and the 151 items which cor-
related highly with their assigned scale scores and manifested low correlations
with the other two scales were included in the final version of the inventory.
This variant of the internal strategy of scale construction—an early forerunner
of the method later used by Jackson (1967) in developing his Personality Re-
search Form—produced scales with intercorrelations ranging from about -.25
(Thinking vs. Social I-E) to around +.25 (Social vs. Emotional I-E).

The Minnesota T-S-E Inventory spawned Drake's (1946) MMPI Social
Introversion (SI) scale, as well as the Thinking Introversion (TI) and Social
Extroversion (SE) scales from the various versions of the Omnibus Personality
Inventory. Drake contrasted the MMPI responses of 50 college females scor-
ing above the 65th percentile on the T-S-E's Social Introversion scale with the
responses of 50 college females scoring below the 35th percentile; all students
were enrolled in a guidance program at the University of Wisconsin. The 70
most discriminating MMPI items were included in the SI scale, which corre-
lated around .70 with the original T-S-E scale in new male and female samples—
about the same value as its correlation with the MMPI Psychasthenia (Pr)
.scale. However, SI correlated around - .80 with the Sociability (Sy) and Social
Presence (Sp) scales from the CPI, and about .90 with Wiggins' Social Malad-
justment (SOC) and Tryon's Introversion (CI-I) item clusters from the MMPI
(Goldberg, 1969).

By the advent of World War II, a revival of American interest in Jung's
theories, especially his conceptions of introversion-extraversion, led to the
development of two rather anomalous I-E inventories: Myers and Briggs's Type
Indicator and Gray and Wheelwright's Psychological Type Questionnaire.
Katherine Briggs and her daughter, Isabel Myers, began the development of
their inventory around 1942, and they constructed numerous revisions over
the next decade. In an attempt to capture more of the flavor of Jung's theo-
retical notions than had been reflected in previous I-E scales, Myers and
Briggs designed their instrument to categorize individuals into dichotomous
types along four interlocking dimensions: Extroversion vs. Introversion (E-I),
Judgment vs. Perception (J-P), Thinking vs. Feeling (T-F), and Sensation vs.
Intuition (S-N). The various quasi-theoretical, quasi-psychometric procedures
used to construct all of the revisions of this inventory are too numerous and
too complex to be detailed here; the interested reader is referred to articles
by Stricker and Ross (1963, 1964a, 1964b). The 1962 version of the Type
Indicator's E-I scale correlated about .65 both with SI from the MMPI and
Sy from the CPI.

During the period of the development of the Type Indicator, another
Jungian team was constructing the Psychological Type Questionnaire (PTQ),
in an analogous attempt to measure three of the four facets in the Jungian
typology (&I vs. E, S vs. N, and T vs. F). In the eleventh revision of the PTQ,
reported by Gray and Wheelwright (1946), 75 items were presented with a
dichotomous response format (e.g., "in giving praise are you [a] reserved,
[b] outspoken"; "Do you [a] spend, [b] save"; "Which do you prefer
[a] keeping house, [b] cooking"). Correlations between the pairs of identi-
cally labeled scales in the two Jungian inventories ranged from .60 for S-N
and T-F to .80 for E-I (Stricker & Ross, 1964b).

During the decade following the initial efforts of the two Jungian teams,
another group was attempting to quantify the psychoanalytic notion of per-
sonality "levels"; the resulting "interpersonal classification system" of Coffey,
Freedman, Leary, and Osorio was later popularized by Leary (1957). As part
of this larger effort, LaForge and Suczek (1955) developed an Interpersonal
Check List (ICL), which was proposed both as an inventory for self-assess-
ment and a schedule for rating others. The fourth revision of the ICL included
134 items, 128 of which were keyed to score two major dimensions (Domin-
ance-Submission [Dom] and Love-Hate [Lov]), and/or alternatively eight
“octant” scores, and/or 16 “category” scores. Each of the 16 categories contained eight items, distributed 1:3:3:1 over four levels of “intensity” or social desirability (e.g., [a] “able to give orders”; [b] “forceful”; [c] “bossy”; [d] “dictatorial”), which were used to provide an overall intensity score (Ain). This sophistication variant of the intuitive and the internal strategies of scale construction yielded a set of octant or category scores whose intercorrelations manifested a pattern approximating that of a circumplex.

About this same time, Eysenck published his Maudsley Personality Inventory (MPI), which included 48 items with a “Yes-No” response format; half of the items were keyed to yield an Extroversion (E) score and the other half were keyed for Neuroticism (N). While Eysenck was originally stimulated by Jung’s I-E conceptualization, it was Hullian learning theory which provided the framework for most of Eysenck’s intensive experimental explorations of the E and N constructs; within Eysenck’s own theoretical structure, extroversion has been equated with “cortical inhibition.” Eysenck’s original E scale was constructed by keying those items which differentiated individuals with high scores from those with low scores on the Rhathymia (carefree impulsiveness) scale from the Guilford’s Inventory of Factors STDCR. Eysenck’s resulting E scale correlated about .80 both with the Guilford’s R scale and with the Social I-E scale from the Minnesota T-S-E Inventory. As already noted, the slight negative correlation between the two MPI scales was “corrected” in the corresponding two scales in the Eysenck Personality Inventory, which was published in 1963.

It is important to realize that even those inventories which do not include any single scale explicitly labeled as “introversion-extroversion” often contain scales which purport to measure individual differences within the same general behavioral class, and factor analyses of these scales often produce factors which are then labeled as I-E. For example, Cattell’s 16PF included measures of Shyness (Sh), Surgency (S), and Dominance (D), all of which loaded highly in a second-order factor which Cattell explicitly called I-E. And Gough’s CPI included measures of Dominance (Do), Sociability (Sy), and Social Presence (Sp), all of which loaded highly on a CPI factor labeled Person Orientation by Nichols and Schnell (1963). The correlation between the factor scores from these two inventories was around .75 (Goldberg, 1969).

MASCULINITY-FEMININITY

While the search for new measures of I-E can be viewed as a response to an unusually compelling theory of individual differences, the quest for new measures of masculinity-femininity (M-F) has had no such persuasive theoretical rationale. Rather, the proliferation of M-F scales over the years can be viewed as a somewhat indolent reaction to the sheer convenience afforded by nature’s provision of two clearly differentiated sexes. The external strategy of scale construction demands some non-test criterion against which to validate items; the fact that males and females abound and that their criterion status is typically obvious appears to have provided all the stimulus needed for the repeated construction of diverse measures of “psychological” M-F.

Moreover, while it could be argued that scales differentiating normal males from females might turn out to be the most potent predictors of homosexuality (Gough, 1952, 1966b), the evidence for this hypothesis is far from compelling. In fact, M-F scales have only rarely been validated against any criterion of homosexuality.

On the other hand, interest in the psychological differences between males and females has a long history in both anthropology and sociology as well as in psychology. Important early reviews of the psychological literature can be found in Woolley (1910, 1914), Hollingworth (1916, 1918), Allen (1927, 1930), and Miles (1935). While a number of investigations of M-F differences were undertaken before World War I, it was the early psychometric work of Terman which provided the most important impetus for the development of later M-F scales. In 1922 Terman began his studies of the responses made by males versus females to diverse stimuli, a project which was to engage his attention for the next fifteen years.

By 1932 Carter had developed the first M-F scale from the SVIB item pool, composed of those items whose responses differentiated males from females in a sample of 114 pairs (including 38 pairs of mixed-sex twins) from grades 7 to 12 (see Carter & Strong, 1933). Within the next few years Strong developed an M-F scale based upon the responses of a college sample (154 pairs of Stanford students, matched on age, college class, aptitude test scores, and college GPA) and another based upon the responses of an adult sample (335 pairs, 277 of which were husband-wife pairs). Strong later constructed three additional M-F scales—each with different scoring weights—based upon the total sample of 603 males and 603 females, and one of these was incorporated into the standard SVIB profile. In contrast to Strong’s use of the external strategy for M-F scale construction, Guilford labeled one of the scales in his 1934 Nebraska Personality Inventory as Masculinity on the grounds that sex had a loading of .84 on that factor; a revised version of the same scale was published in 1943 as factor M in Guilford and Martin’s Inventory of Factors GAMIN.

In 1936 Terman and Miles published their landmark volume, Sex and personality, which reported the results of a series of investigations of sex differences. Diverse types of stimulus content were administered to various male and female samples, including (a) word associations (a stimulus word followed by four alternative responses with instructions to select the alternative most highly related to the stimulus word); (b) inkbolt associations (an inkbolt followed by four alternative responses with similar instructions); (c) information items (e.g., “the length of a brick is: 6”, “8”, “10”, “12”); (d) emotional or ethical situations, with directions to select one of four alternative emotional reactions; (e) interest items including occupational titles, types of people, avocational activities, books, drawing, reporting, and sightseeing preferences—all presented with a “Like-Dislike-Neutral” response format; and (f) I-E and other personality, attitude, and opinion items presented with a “Yes-No” response format. From those items whose responses differentiated male from female samples, Terman and Miles (1936) selected 456 items for Form A and 454 other items for Form B of their Attitude-Interest Analysis Test (AIAT),

XIV. PERSONALITY SCALES AND INVENTORIES

325
grouped to form seven scales (Word Association, Inkblot Association, Information, Emotional-Ethical responses, Interests, Personality and Opinions, and I-E responses). The scales varied in length from 18 items (Inkblot Association) to 119 items (Interests), and all of the items were keyed in the masculine direction. Scale reliabilities ranged from about .25 (I-E responses) to about .90 (Personality and Opinions), and intercorrelations among the scales ranged from a low of .15 (Word Association vs. Emotional-Ethical responses) to a high of almost .50 (Information vs. Personality and Opinions), all in single-sex samples. The reliabilities and scale intercorrelations were higher in mixed-sex samples, and the AIAT total score, as well as the Interest subscore, provided excellent discrimination between male and female samples.

One of Terman’s students, E. Lowell Kelly, administered the AIAT to 134 male homosexuals, 46 of whom were classified as taking an active sexual role, 77 as taking a passive role, and 11 whose sex role was less clear. The distribution of AIAT scores for the passive homosexuals lay midway between the distributions for the normal male and female samples, while the distribution of scores for the active homosexuals was similar to, though somewhat more masculine than, the distribution for normal males. In general, the AIAT turned out to make rather poor discriminations between the normal male and the total homosexual samples. Consequently, Kelly devised an Invert scale composed of those items whose responses differentiated the passive homosexual sample from normal males; the Invert scale had a rather low correlation with the original AIAT score.

Kelly’s research, which was reported in the Terman and Miles (1936) volume, may have led Hathaway and McKinley to attempt to construct a predictor of homosexuality for the MMPI. Their Mf scale included MMPI items which differentiated a sample of 13 homosexual males from a normal sample, as well as items with high correlations with the Invert scale, plus some items whose responses differentiated normal male from female samples. While the original Mf scale was developed solely as an experimental measure, the scale soon got locked into the standard MMPI profile package, and it has never been revised (Hathaway, 1956).

During the decade following the end of World War II, a number of inventory developers—perhaps stimulated by the Mf scales from the MMPI and the SVIB—included new Mf scales in their own instruments. In the 1946 revision of his Preference Record—Vocational (PR—V), Kuder included an optional Mf scale, which was dropped from the 1956 revision. In 1947 Bell published the Personal Preference Inventory (PPI), a supplement to his older Adjustment Inventory; the PPI included 90 items with a “Yes-No” response format, 30 scored on each of three scales (Mf, Criticalness, and Perceived Economic Status). By 1952 Gough had developed his Femininity (F) scale, composed of those 58 items (out of over 500 initially investigated) whose responses differentiated most highly between male and female samples. The scores from a subset of 32 F items administered to 38 homosexual and 38 heterosexual prison inmates significantly differentiated between the two samples (Gough, 1952). The longer F scale was included in Gough’s California Psychological Inventory, which was published in 1956.

In the past 15 years, Mf has become the single most popular construct for inclusion in new personality inventories. Mf scales have been constructed for Jenkins’ How Well Do You Know Your Interests?, Welsh’s Figure Preference Test, Holland’s Vocational Preference Inventory, Heist and Yonge’s Omnibus Personality Inventory, Fricke’s Opinion, Attitude, and Interest Survey, Comrey’s Personality Scales, and Giddan, King, and Lovell’s Academic Behavior Inventory. As already noted, there is some significant item overlap between these various inventories, with the result that a number of these different Mf scales share a substantial proportion of common items.

At least ten investigators have correlated the Mf scales from various inventories, including: (a) Heston (1940), who correlated Mf scales from the SVIB, MMPI, and Kuder’s PR—V; (b) de Cillis and Orbison (1950), who used the AIAT and the MMPI; (c) Shepler (1951), who included the AIAT, SVIB, MMPI, and a projective measure; (d) Gough (1957), who correlated CPI scales with those from the MMPI and SVIB; (e) Stanek (1959), who returned to the AIAT and MMPI; (f) Barrows and Zuckerman (1960), who included the SVIB, MMPI, and the Guilford—Zimmerman Temperament Survey (GZTS); (g) Nichols (1962), who correlated Mf scales from the MMPI, CPI, and Guilford—Martin Gamin; (h) Engel (1966), who used the AIAT, SVIB, MMPI, CPI, and a projective measure; (i) Klopfner (1966), who included only the SVIB and MMPI; and (j) Himelein and Stoup (1967), who used the SVIB, MMPI, and GZTS. A summary of the major results from these studies is presented in Table 2.

As Table 2 indicates, the correlations among Mf scales are moderated by a number of different variables. First of all, Mf scale intercorrelations are markedly higher in mixed-sex samples than in single-sex samples. Moreover, as Gough (1957) and Strong (1943) have demonstrated, these correlations are also significantly affected by the age of the subjects (e.g., high school vs. college samples). In addition, as Terman and Miles (1936) originally noted, the correlations are higher among Mf scales composed of items with similar content than between scales of differing content, and they are clearly highest when the scales share a set of common items (e.g., the CPI and the MMPI). Furthermore, Mf scale correlations with biological sex are moderated by the strategy used in scale construction (Nichols, 1962; Goldberg, 1970); scales composed of items which are both “obvious” sex discriminators (e.g., “I like to wear pretty, frilly panties”) and which empirically differentiate between the two sexes provide more clear separation between male and female samples than do Mf scales composed of either more subtle discriminating items or of “obvious” items not subjected to any empirical test. These results, which are in accord with those from studies by Goldberg and Slovic (1967) and Norman (1963) on the relationship between face and predictive validity, suggest the substitution for post Mf scales of the following single item: “I am a male: True or False.”

**PERSONAL VALUES AND MANIFEST NEEDS**

In contrast to the spirit of ruthless empiricism which has characterized the quest for Mf measures, quite a few personality inventories have been focused on two theories of individual differences—the first proposed by the philosopher
Table 2
Intercorrelations among Measures of Masculinity-Femininity

I. Correlations Computed within Single-Sex Samples: Male Samples above the Diagonal and Female Samples below

<table>
<thead>
<tr>
<th></th>
<th>MMPI</th>
<th>SVIB</th>
<th>CPI</th>
<th>AIAT</th>
<th>GZTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMPI</td>
<td>-</td>
<td>.53^1</td>
<td>.43^d</td>
<td>.30^b</td>
<td>.31^h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.50^c</td>
<td>.39^d</td>
<td>.66^e</td>
<td>.28^1</td>
</tr>
<tr>
<td>SVIB</td>
<td>.48^k</td>
<td>-</td>
<td>.41^d</td>
<td>.56^c</td>
<td>.24^h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.55^c</td>
<td>-</td>
<td>.20^1</td>
<td></td>
</tr>
<tr>
<td>CPI</td>
<td>.52^1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AIAT</td>
<td>.17^g</td>
<td>.67^c</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

II. Correlations Computed within Mixed-Sex Samples

<table>
<thead>
<tr>
<th></th>
<th>MMPI</th>
<th>SVIB</th>
<th>CPI</th>
<th>PR-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVIB</td>
<td>.89^g</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.69^a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPI</td>
<td>.71^j</td>
<td>.59^d</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>PR-V</td>
<td>.68^a</td>
<td>.73^a</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>AIAT</td>
<td>.65^j</td>
<td>.69^d</td>
<td>.60^f</td>
<td></td>
</tr>
<tr>
<td>SEX</td>
<td>.74^a</td>
<td>.63^a</td>
<td>.65^e</td>
<td>.61^a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.78^e</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^aHeston (1948): N = 79 mixed sex.
^bde Cillis and Orbilson (1950): N = 129 males; N = 50 females.
^cShepley (1951): N = 57 males; N = 67 females.
^dGough (1957): N = 152 males.
^eGough (1957): N = 7,628 mixed sex (high school).
^fGough (1957): N = 1,590 mixed sex (college).
^gStanek (1959): N = 132 females.
^hBarrows and Zuckererman (1960): N = 2,296 males.
^iNichols (1962): N = 100 males; N = 100 females.
^lHilmstein and Stoup (1967): N = 60 males.

Edouard Spranger and the second by the psychologist Henry Murray. However, it was probably the popularity of the initial inventories developed under each of these two frameworks rather than the theories themselves which stimulated so many psychometricians to return to these same traits, since there have been few attempts to justify a focus on these particular individual differences beyond those propounded by the original theorists.

The English translation of Spranger’s (1928) book, Types of men, which posited six major value orientations (represented by the theoretic, economic, aesthetic, social, political, and religious man) led Vernon and Allport (1931) to construct their seminal Study of Values. From Spranger’s writings, Vernon and Allport selected a set of statements presumably descriptive of each of the six value types, and they wrote items which contrasted two or more of these values. The 45 items whose alternatives had nearly equal popularity (endorsement frequency) in college samples and whose responses related most highly to an a priori keying of the six scales were included in the original version of the Study. While the ipsative character of the Study’s scoring procedures constrained each subject’s mean score across all six scales to the same value, the reliability (homogeneity) values for the six scales differed greatly; they ranged from around .50 (Social and Political) to about .85 (Aesthetic and Religious). A revised version of the Study, published in 1951 by Allport, Vernon, and Lindzey, also contained 45 items, but the scale reliability values had been raised to the .75 to .90 range. A third edition by the same authors, published in 1960, was virtually identical to the 1951 model.

In contrast to the intuitive-internal strategy of scale construction used by Allport, Vernon, and Lindzey, the Interest-Values Inventory (IVI), which was published by Maller and Glaser in 1939, was constructed by the external strategy (see Glaser & Maller, 1940). Fifty college students in each of four broad curricular fields (mathematics and science, arts and music, social work and nursing, and business and advertising) were used as criterion groups to develop scales purportedly measuring the relative dominance of four of Spranger’s six basic values (theoretic, aesthetic, social, and economic). The IVI contained 116 items of diverse types, including 10 sets of 4 stimulus words with instructions to select the most preferred word, 10 association items with instructions to select the most highly related of 4 alternative responses, 48 answers to 12 basic questions each with a Like-Neutral-Dislike response format, and 48 trait-descriptive adjectives with instructions to indicate whether each trait had “Strong-Average-Weak” self-applicability. While the Study of Values has joined the SVIB as one of the two oldest inventories in popular use today, the IVI has a relatively short history and is now out of print.

Spranger’s theory, which spawned these two inventories, had only limited progeny in comparison with the theoretical views of Henry Murray, whose offspring already include Edwards’ Personal Preference Schedule, Stern’s Activities Index, Jackson’s Personality Research Form, Hellbrun’s need scales for the Adjective Check List, and Hase’s need scales for the CPI; moreover, there are enough questions concerning the paternity of Guilford, Christensen, and Bond’s DF Opinion Survey to again implicate that same prolific theory. For a framework described as no more than a “rough, preliminary plan [Murray et al., 1938, p. 143],” this is quite a record!

Henry Murray and his colleagues at the Harvard Psychological Clinic were highly influenced by psychoanalytic theory, and the anatomy of personality structure described in their classic 1938 volume, Explorations in personality, represented a unique amalgamation of the thinking of Freud and his followers with that of Allport and other academic psychologists. The major individual
XIV. PERSONALITY SCALES AND INVENTORIES

The Personality Variables Posited by Murray et al. (1938)

<table>
<thead>
<tr>
<th>Major Manifest Needs (20)</th>
<th>Trait Descriptive Terms</th>
<th>Original Type</th>
<th>Measure Included in Later Inventories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnewment</td>
<td>Submissive, acquiescent, passive</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Achievement</td>
<td>Ambitious, competitive, striving</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Affiliation</td>
<td>Accepting, cooperative, affiliating</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Aggression</td>
<td>Authoritative, critical, severe</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Autonomy</td>
<td>Independent, defiant, stubborn</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Counterfactual</td>
<td>Resultive, determined, adventurous</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Confusion</td>
<td>Defensive, respectful, compliant</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Dependence</td>
<td>Self-defensive, self-destructive</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Dominance</td>
<td>Aggressive, forceful, decisive</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Exhibition</td>
<td>Dramatic, conspicuous</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Harmavoids</td>
<td>Fearful, timid, cautious, careful</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Inflexible</td>
<td>Sensible, shy, reserved</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Nurturance</td>
<td>Sympathetic, gentle, protective</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Order</td>
<td>Organized, clean, neat, precise</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Play</td>
<td>Playful, amusing, jolly</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Reaction</td>
<td>Excessive, aloof, discriminating</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Sentience</td>
<td>Sensible, sensitive, aesthetic</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Sex</td>
<td>Erotic, sexual, seductive</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Success</td>
<td>Dependable, helpful, informed</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
<tr>
<td>Understanding</td>
<td>Intellectual, creative, logical</td>
<td>10*</td>
<td>EPERS, ACL, CPI, AI, PRF</td>
</tr>
</tbody>
</table>

Other Manifest Needs (7)

Acquiescence
Harmavoids
Inflexible
Nurturance
Order
Play
Reaction
Sentience
Success
Understanding

General Traits (12)

Average (Harmavoids + Inflexible + Nurturance)
Creativity
Conscientiousness/Discreetness (Coordination of action and thought)
Efficiency vs. Productivity
Endurance
Excentricity/Endarrestion (Outward vs. towards orientation)
Extraversion/Introversion (Imaginative vs. practical)
Impulsively/Deliberately (Immediate vs. delay)
Integrity
Projection/Objectivity (Projection vs. detachment)
Radical/Conservative
Sameness/Change

Other Internal Factors (6)

Ego Ideal (Level of achievement and level of aspiration)
Narcissism (Self-love)
Superego Integration (Acceptance of one's conscience)
Superego Conflict (Guilt and remorse)

Latent Needs (6)

Repressed Abnewment (Positivity and Measurability)
Repressed Aggression (Fate and Sudden)
Repressed Criticism (Worshipful)
Repressed Dominance (Omnipotence)
Repressed Exhibitionism (Exhibitionist)
Repressed Heterosexuality
Repressed Homosexuality
Repressed Sobriety (Helplessness)

Microbiothermal Variables (5)

Expansion/Constriction
Superiority/Inferiority
Optimism/Pessimism
Social Solidarity
Neurosis

* Fifteen additional attitude items included in the Sentiment Questionnaire.

The first major inventory developed to compete with the original Harvard questionnaire was Edwards' Personal Preference Schedule (EPERS), which was introduced in 1953, fifteen years after the publication of Explorations in personality. The EPERS included 225 items keyeed for 15 scales (see Table 3), each item a forced-choice between two statements which had been roughly matched on social desirability, one from each of two scales from the original Harvard questionnaire. Reliability (homogeneity) coefficients ranged from about .60 (Deference, Exhibition) to around .85 (Heterosexuality, Abasement, Aggression). Since the EPERS scales were ipatively scored, scale intercorrelations were quite low; the highest correlation, that between Affiliation and Nurturance, was around .45 (Edwards, 1954). By 1958, five years after the publication of the EPERS and the initial development of Gough's 300-item Adjective Check List (ACL), Heilbrun had constructed a set of ACL scales purporting to tap the same 15 traits included in the EPERS. Heilbrun's 15 need scales, plus 2 he constructed as measures of Defensiveness and Counseling Readiness, were added to 7 ACL scales previously developed by Gough and published in Gough and Heilbrun's (1965) manual for the ACL. Correlations among the ACL need scales ranged into the .70s (Order vs. Endurance) in a sample of 800 subjects (Gough & Heilbrun, 1965). About the same time as the ACL manual was published, Hase constructed 11 need scales from the CPI item pool for a larger project investigating the comparative validity of various strategies of scale construction (Hase & Goldberg, 1967). Eight of these recent CPI scales purportedly measure the same traits as those included among the 15 EPERS and ACL scales (see Table 3).
“presses” along with measures of personal “needs” (see Stern, 1970). Stern’s various Environmental Indexes included scales directed at the measurement of the “climates” in high schools, colleges, evening colleges, the Peace Corps, and other organizations. The parallel AI, which included 300 items with a like-dislike response format now provides 49 scores, including 30 need scores (see Table 3), 12 factor scores, 4 second-order factor scores, plus validity and academic aptitude indices. Of the eight AI “needs” labeled differently than any included in the Murray list (Practicalness-Impracticalness, Adaptability-Defensiveness, Ego Achievement, Reflectiveness, Energy-Passivity, Fantasied Achievement, Humanities and Social Science, and Science), the first four doubtless correspond to Extroversion-Introversion, Defendence, Ego Ideal, and Exocathexis-Endocathexis, while the fifth may be an analogue of Succorance.

Another recent and unusually sophisticated attempt to measure some of these same constructs can be found in Jackson’s (1967) Personality Research Form (PRF). Use of the two PRF forms permits the measurement of 19 constructs (plus Cognitive Structure, a measure of thinking rigidity and intolerance of ambiguity)—a few more scales than the 15 from the EPPS and ACL, but less than the 30 provided in the AI. The PRF scales included those items from a much larger set which satisfied both intuitive and internal criteria for scale membership and which manifested relatively low correlations with other scale scores and with a putative measure of social desirability response set. While PRF scale reliability coefficients were generally quite high, some of the scale intercorrelations ranged up to almost .65 (Jackson, 1967).

While the EPPS, ACL, AI, and PRF scales were directly modeled after the Murray system, one other inventory can be viewed as an indirect product of the original Harvard questionnaire. Over the years, Guilford and his associates have extensively analyzed a large set of older personality and interest statements and later devised new items as more direct measures of the factors which emerged from the earlier analyses. Guilford, Christensen, and Bond’s DF Opinion Survey (DFOS), which was published in 1954, included 300 new and old items, 30 for each of 10 scales (Need for Attention, Liking for Thinking, Adventure, Self-Reliance, Aesthetic Appreciation, Cultural Conformity, Need for Freedom, Realistic Thinking, Need for Precision, and Need for Diversion). The internal strategy of scale construction used to fashion this inventory produced scales with reliability (homogeneity) values ranging from about .65 to .95. Since correlations between the DFOS scales and those from the EPPS, ACL, AI, and PRF have not been reported, it is difficult to assess the extent to which the concepts of the Guilford group have diverged from those originally devised by the Harvard team.

OTHER TRAITS

The constructs classified in the first six columns of Table 1 certainly do not exhaust the domain of individual differences targeted by personality scales and inventories, and only space limitations prevent the inclusion of other salient assessment foci. For example, attempts to measure aspects of liberal vs. conservative ideology date back to the 1920s (see Shaw & Wright, 1967), culminating in such significant measurement milestones as Lentz’s (1930) C-R (Conservatism-Radicalism) Opinionnaire, developed in the late 1920s and published in 1935; Rundquist and Sletto’s (1936) Economic Conservatism Scale; Levinson’s (1949) E (Ethnocentrism) Scale and the California F (Fascism) Scale, both given wide currency via the classic work on the authoritarian personality (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950); and Rokeach’s (1960) Dogmatism Scale. More recently, a flurry of research activity has centered on Rotter’s (1954) distinction between internal vs. external locus of control (internals perceiving events as manipulable by oneself, externals perceiving events as determined by fate, chance, or the manipulation of others). The instrumentation for measuring this construct was developed in doctoral dissertations at Ohio State by Phares in 1955 and by James in 1957; for reviews of the research stimulated by these measures, see Lefcourt (1966) and Rotter (1966). For details of an inventory developed to assess other aspects of Rotter’s (1954) social learning theory, see Liverant (1958).

During the 1940s, Thurstone redirected his attention from attitude scaling to personality structure. After factoring items from Guilford’s various batteries, Thurstone (1951) concluded that seven factors were worth developing further, and he tried out a host of items from previous scales and inventories as measures of each factor. His resulting Temperament Schedule, which was published in 1949 and revised in 1953, included seven scales labeled Active, Vigorous, Impulsive, Dominant, Stable, Sociable, and Reflective. A new inventory, Comrey’s Personality Scales, has been constructed via the same general strategy. Over the years, Comrey and his collaborators have been clustering sets of new and old personality items to form “factored homogeneous item dimensions,” which they have used in a series of factor-analytic investigations (e.g., Comrey, 1961, 1962, 1964; Comrey & Jackson, 1966). Eight of the resulting factor scales (Trust, Orderliness, Conformity, Activity, Neuroticism, Empathy, F-E, and M-F) were included in the new Comrey inventory, which was published in 1970.

While both Thurstone and Comrey relied upon traditional verbal items, two other inventory developers have avoided these standard types of questions completely. Welsh devised a figure preference test for his doctoral dissertation at Minnesota in 1949; 400 line-drawings were used as items with a “like-dislike” response format. A number of scales were constructed by the external strategy with such criterion groups as males vs. females, psychiatric patients vs. normal subjects, children vs. adults, and artists vs. non-artists—the latter used to construct the popular Barron-Welsh Art Scale. These scales were supplemented by others focused upon preferences for specific kinds of designs (e.g., freehand vs. ruled lines) and published in 1959 as Welsh’s Figure...
Preference Test (WFPT). In contrast, Holland used a set of 300 occupational titles (with the same like-dislike response format) to construct his Vocational Preference Inventory (VPI), which was first published in 1953. Holland used the intuitive strategy to select and group items, though he later eliminated some scales and refined others by means of internal consistency analyses (see Holland, 1958). Nine original content scales (Intelectuality, Social Responsibility, Conformity, Verbal Activity, Emotionality, Control, Aggressiveness, Status, and M-F) were revised and relabeled in the 1965 revision of the VPI to furnish new scales called Realistic, Intellectual, Social, Conventional, Enterprising, Artistic, Self-control, Status, and Masculinity.

While responses to the items on the WFPT and the VPI are not easily construed as being more or less desirable, most other personality scales and inventories have contained verbal items whose responses were rather obviously related to their overall favorability. Over the years, inventory developers have been concerned about potential individual differences in impression management, and as early as 1931 Vernon and Allport attempted to control for such a possibility by using a forced-choice item format in their Study of Values. More recently, Allen Edwards (1957) redirected psychometric attention to this problem, which he relabeled “social desirability response set,” and his EPSS employed a forced-choice format—two statements of approximately equal desirability administered with instructions to select that member of the pair which is most self-descriptive. Soon, a number of other inventory constructors adopted the same general tactic. Gordon used an item format popular in the vocational interest domain—item triads (roughly equated on desirability) with instructions to rank order the three items—to construct his Survey of Interpersonal Values (SIV), which was published in 1960. Thirty item triads were included in the SIV, keyed for each of six scales (Support, Conformity, Recognition, Independence, Benevolence, and Leadership). In a similar vein, Thordike arranged items of approximately equal desirability in sets of ten, administered with instructions to select the three most characteristic and the three least characteristic items within each set. Thordike’s resulting inventory, the Dimensions of Temperament (DOT), was published in 1963. The DOT contained ten scales labeled Sociable, Ascendant, Cheerful, Placid, Accepting, Tough-minded, Reflective, Impulsive, Active, and Responsible. Both the SIV and the DOT—like the Study of Values and the EPSS—provided ipsative scores.

Not all recent inventories, however, have employed forced-choice and/or ipsative procedures, and a number of these have provided far more scales than those included in most earlier inventories. Gryger attempted to measure 33 psychoanalytic constructs (e.g., orality, hoarding, narcissism) by modifying Krout and Tabin’s (1954) Personal Preference Scale, one of the first psychoanalytically-oriented personality inventories. Gryger’s 325-item Dynamic Personality Inventory was published in 1961. Even more scores were provided in Cattell and Horn’s Motivation Analysis Test (MAT), which was first published in 1959 and revised in 1969. While both inventories share a quasi-dynamic flavor, the MAT was not directly modeled after psychoanalytic theory; instead, the constructs were derived from a series of factor-analytic studies of motivation items. The MAT provided four types of scores (Integrated, Unintegrated, Total, and Conflict) for each of five “ergs” (Mating, Assertiveness, Fear, Narcissism, and Puqacity) and for each of five “sentiments” (Superego, Self, Career, Home, and Sweetheart), plus five total scores (Integration, Interest, Conflict, Autism, and Information). For further details of this unusual inventory, see Cattell, Radcliffe, and Sweny (1963).

The proliferation of scales within one inventory may have reached its apex, however, with the publication of Edwards’ (1966) Personality Inventory (EPI), which included 53 scales. Edwards developed an initial pool of over five thousand personality descriptive statements, gleaned from informal conversations with individuals who were asked to describe others they knew. Elimination of duplicated statements left a pool of 2,824 items which were grouped intuitively, administered to a college sample, and then subjected to a series of factor analyses (Edwards, 1966). About 1,200 of the items were finally used to measure the 53 EPI scales, while another 300 items provided parallel measures of 14 scales. While the EPI is administered with instructions to “predict how people who know you best would mark each statement if they were asked to describe you,” Edwards (1969) has shown that there is essentially no difference between EPI scores based upon these novel instructions and scores based upon traditional self-descriptions. On the other hand, most of the EPI items—like those included in Jackson’s (1967) PRF—are less obviously related to overall response desirability than were many of the items from previous inventories.

SOME CONCLUDING REMARKS

If the future is anything like the past, new personality scales and inventories are at least as likely to be focused upon constructs arising out of applied societal pressures as upon any theories of personality. In fact, the most potent source of variance in the determination of the constructs for past scales and inventories has been sheer historical accident. For better or for worse, psychologists have tended to measure those constructs already identified by their predecessors, and in addition they have tended to borrow heavily from past item pools. Items devised around the turn of the century may have worked their way via Woodworth’s Personal Data Sheet, to Thurstone and Thurstone’s Personality Schedule, hence to Bernreuter’s Personality Inventory, and later to the Minnesota Multiphasic Personality Inventory, where they were borrowed for the California Personality Inventory and then injected into the Omnibus Personality Inventory—only to serve as a source of items for the new Academic Behavior Inventory. As a result of the widespread practice of item borrowing, there is substantial item overlap between a number of present inventories (one result of which is that convergent validity coefficients computed between scales from two inventories—generally lamented as being too low—may, in fact, be spuriously high).

Moreover, among those inventory developers who have avoided past constructs or past item pools, another trend is equally clear. Each original trait has been gradually bifurcated into smaller and smaller constructs. As an
example, Introversion-Extroversion was later divided into three components, one of which was social introversion; the latter, in turn, has been fractionated into at least five components, one of which was dominance; and recently, dominance has shattered into 30 to 40 "facets" (Butt & Fiske, 1968). Analogously, anxiety (Taylor, 1953) has been dichotomized into test anxiety and general anxiety, and the latter, which was construed as five independent factors in the 16PF, has more recently exploded into myriads of "anxiety-by-situation interactions" (Endler, Hunt, & Rosenstein, 1962). Adjustment was once just that—a single global construct; over the years, the construct has been shredded so finely that Jackson and Messick's new Differential Personality Inventory purports to measure some 28 varieties of maladjustment.

In some sense, scientists invariably proceed in the dark, and only their past activities are illuminated. However, historians have long held that the reflected glow from a carefully lighted past may help to outline at least some parts of the near future. Hopefully, this brief history of personality scales and inventories may provide enough light to keep future psychometric investigators from stumbling over the same obstacles that seem to have blocked their progress in the past.
References


Bernreuter, R. G. The measurement of self-sufficiency. Journal of Abnormal and Social Psychology, 1933, 28, 291-300. (a)

Bernreuter, R. G. The theory and construction of the personality inventory. Journal of Social Psychology, 1933, 4, 387-405. (b)


Goldberg, L. R. Diagnosticians vs. diagnostic signs: The diagnosis of psychosis vs. neurosis from the MMPI. Psychological Monographs, 1965, 79 (9, Whole No. 602).


Gough, H. G. Graduation from high school as predicted from the California Psychological Inventory. *Psychology in the Schools*, 1966, 3, 208-215. (a)


Guilford, J. P., & Guilford, R. B. Personality factors N and CD. *Journal of Abnormal and Social Psychology*, 1939, 34, 239-248. (b)


Hollingworth, L. S. Comparison of the sexes in mental traits. Psychological Bulletin, 1918, 15, 427-432.


Laird, D. A. A mental hygiene and vocational test. **Journal of Educational Psychology**, 1925, 16, 419-422. (b)


Pressey, S. L. A group scale for investigating the emotions. *Journal of Abnormal and Social Psychology*, 1921, 16, 55-64.

Pressey, S. L., & Chambers, O. R. First revision of a group scale designed for investigating the emotions, with tentative norms. *Journal of Applied Psychology*, 1920, 4, 97-104.


Spranger, E. *Types of Men (Lebensformen)*. Halle: Niemeyer, 1928.


Williamson, E. G., & Darley, J. G. The measurement of social attitudes of college students: II. Validation of two attitude tests. *Journal of Social Psychology*, 1937, 8, 231-242. (b)

