OBJECTIVE DIAGNOSTIC
TESTS AND MEASURES

Lewis R. Goldberg
University of Oregon and Oregon Research Institute, Eugene, Oregon

In a previous chapter in the Annual Review of Psychology covering a quite different topic, Tulving & Madigan succumbed to the torment of the Annual Review author and wished aloud that "... at least some writers, faced with the decision of whether to publish or perish, should have seriously considered the latter alternative" (149, p. 442). If this final solution to the problems faced by reviewers seems at all unappreciative, consider the following dilemma: (a) the present chapter, the first on a topic that will now be reviewed once every 3 years, must be held to approximately 25 printed pages; (b) in 25 Annual Review pages, if one avoided all text whatsoever, one could include a bibliography of about 675 titles; (c) during the 3 year period 1970–1972, the period chosen for this review, there appeared over 750 titles on one instrument, the MMPI, alone (12), and thus it would not be possible even to list merely a bibliography of MMPI reports; (d) each year there are over 3000 titles (including books, chapters in edited books, and journal articles) of some relevance to this topic (12), or close to 10,000 for the entire period. So, clearly, this review must be ruthlessly selective, hacking away most of this bloat with machete-like abandon.

Towards this end, efforts have been made to avoid duplicating the material covered in chapters on assessment, human abilities, personality, attitudes and opinions, personnel selection, scaling, and test theory in previous Annual Review volumes, and in chapters on personality, genetics of behavior, social and cultural

1The author wishes to acknowledge the contributions of Endel Tulving and Stephen Madigan (149), who, while never having met the author nor having seen a draft of this chapter, have provided quite substantial guidance by previous example. Any errors in the manuscript, therefore, are clearly their fault. Kudos, on the other hand, might be sent to the author, who is always most appreciative. As he is for the financial support which permitted him to undertake this adventure, support provided by Grant MH-12972 from the National Institute of Mental Health, U.S. Public Health Service. Investigators who feel neglected or misunderstood are encouraged to seek relief from the Public Health Service.
influences on psychopathology, human information processing, and statistics in the present volume. However, any such review shares a problem with the zone defense in professional football: namely that excellent scientific reports, like talented tight ends, can often slip through the cracks between the zones. Readers who feel that their favorite instrument was given short shrift, or no shrift, are encouraged to write the author, who may be able to provide them with an extensive specialized bibliography.

In the current Annual Review of Psychology master plan, this chapter is part of the genus Abnormal and Clinical Psychology, and the species Assessment of Psychopathology. The “objective” in the title is meant to contrast with “projective” (whatever that may convey), not necessarily with “subjective,” as one might have guessed. The chapter itself is arranged as a diagnostic instrument. Since the material becomes increasingly more technical as one proceeds through it, one can assess his degree of involvement with the field simply by noting the page on which he moves to potentially greener pastures.

FOR THOSE WHO HAVE WANDERED INTO THIS CHAPTER FROM ACROSS THE BORDERS . . .

If you read only one book in this field every 10 years, the choice for you has now been made easy: read Jerry Wiggins' Personality and Prediction: Principles of Personality Assessment (155). A magisterial integration of the concepts and the research from a field whose complex sprawl had easily bested all previous attempts, this textbook is the most important single volume written about assessment since 1957 and the original edition of Cronbach & Gleser's Psychological Tests and Personnel Decisions. From the opening chapter on correlation and multiple regression to the penetrating analyses of five milestone assessment projects at the end, Wiggins presents an even-handed account of the conflicts and convergences of this inchoate field. Along the way, the reader will encounter such topics as moderator variables and the prediction of predictability, contingency tables and actuarial prediction, clinical prediction and its automation, decision theory and personnel assignment, observational techniques (including reliability theory, trait attribution, and behavioral analysis), structured techniques, and the role of theoretical models in assessment science and practice. Buy it; you'll like it.

At a considerably more technical level, one other book published during this period stands out as a signal achievement. Cronbach & Gleser (23) have returned again, now with two collaborators, Nanda & Rajaratnam, to produce a second methodological landmark. Their multifaceted generalization of classical reliability theory, entitled The Dependability of Behavioral Measurements (24), is an evolutionary reconceptualization of the fundamental measurement base for objective diagnostic procedures. Warning: This is not a book for the mathematically fragile or the conceptually timid. Translation: I found it hard sledding.

Not so with two excellent compendia from the period, McReynolds' Advances in Psychological Assessment, Volume 2 (94) and Butcher's Objective Personality Assessment (13). The former includes chapters on the assessment of places, of
children, of managerial talent, of aggression, of intrinsic motivation, of therapy outcomes, and of psychotic behavior, plus a mercilessly long mother by the present author surveying the history of personality scales and inventories (54). The latter book, which focuses primarily on the MMPI, includes two absolutely superb chapters, one by Paul Meehl (98) and the other by Warren Norman (107). Meehl has returned to writing about assessment again, albeit in some obscure places (96, 97, 99, 100). His writing style has changed over the years since that bodacious flurry of extraordinarily influential papers published between 1945 and 1960, and his papers now read as exquisite rambles through a tortuously labyrinthine phenomenology. As a consequence, one leaves Meehl’s writing today with a sense of awe, rather than, as before, with a new point of view. Norman (107) is something else again. If Wiggins wins the literary prize for best novel-length work, Norman is the clear winner in the short-story-length class. Word for word, his Psychometric Considerations for a Revision of the MMPI contains more measurement wisdom than anything else published during the decade.

This was the period of the revised test handbook, with five old friends of the practitioner-on-the-street spawning new ones. These mighty tomes, all rich sources of material on the merits and demerits of their respective tests, include Matarazzo’s comprehensively revised WAIS handbook, Wechsler’s Measurement and Appraisal of Adult Intelligence (91); Campbell’s Handbook for the Strong Vocational Interest Blank (14); the first volume of Dahlstrom, Welsh & Dahlstrom’s two-volume revision of their MMPI Handbook (26); Megargee’s California Psychological Inventory Handbook (101); and Cattell, Eber & Tatsuoka’s Handbook for the Sixteen Personality Factor Questionnaire (18). All five volumes are real handbooks in today’s sense of that term, which means that none of them is small enough to be held with one hand. Though, to be fair, they look puny when compared to the gargantuan products of O. K. Buros (11, 12). His two latest works, which might logically be called handbooks since neither is small enough to be held with both hands, include the monumental Seventh Mental Measurements Yearbook (covering the period 1964–1970 and 1157 tests) and Personality Tests and Reviews (which, among many other significant things, reprints all reviews of personality measures from the first six Yearbooks). Most psychologists know that if they need information about a particular test, they should consult Buros’ latest Yearbook. Less widely known is the fact that virtually all books about assessment are listed there, each followed by lengthy excerpts from most of the reviews these books received in professional journals. For information about published tests and related books, there is simply no better source.

The key word in that last sentence was “published,” obviously only the tip of the assessment ice. Tailor-made instruments are often used in psychological research, and reports of their psychometric properties are sometimes included in journal articles. No one systematically reviews that stuff (other than Psychological Abstracts, which catalogs everything—and, thus, nothing). Now there is such a survey, at least for measures used with children: Johnson & Bommarito’s Tests and Measurements in Child Development (71) includes 309 instruments, well classified and well described. If you wandered in here from developmental psychology, you’ll want to order this one at the same time you order the Wiggins volume.
FOR THOSE WHOSE INTEREST IN THIS CHAPTER IS AT LEAST PARTIALLY PRACTICAL OR APPLIED . . .

If you are into the practice of psychodiagnosis these days, many authors will teach you how to do it right (meaning their way), while many more will urge you to jump ship and give up the enterprise altogether. By and large, the former write or edit books, while the latter write chapters and journal articles. Exceptions to this general law are Raymond B. Cattell and Hans J. Eysenck, who do both things—and often. Eysenck alone had 60 publications during this 3-year period, including 6 books (actually 8, since one edited work was in three volumes [43–45]). For Eysenck’s approach to diagnosis, see (40, 42). For Cattell’s position, see (16). For the views of others, try another chapter or two in the edited book by Mahrer (88)—particularly that of Lorr (82)—or pick up the short booklet by Dahlstrom (25).

Of those critics who fight to make you switch, one of the best is Albee (1). Such pugnacious types, in turn, are opposed by some highly muscular hardhats with a heavy biological emphasis, among them Meehl (97, 99), Rimland (125), Rosenthal (128), and Gottesman & Shields (61). If there’s anyone around who hasn’t yet read Rimland’s audacious argument against psychogenesis, run—do not walk—to the nearest library and start reading it. Rimland is doubtless wrong, as are we all, but after you have read his piece, you will never think quite the same way again. Which means that you may be ready to compromise. If so, read Begelman (5), Gough (62), and Draguns & Phillips (33).

Pecking away at people who use traditional diagnostic syndromes has been a preoccupation among psychologists with a behavior-modification or social-learning orientation towards psychochange (120). Much of the time spent in this business has been devoted to exhausting such hobgoblins as traits, diagnoses, medical models, and similar demons. Apparently most of this dangerous work has now been accomplished, and some authors can turn their attention to other pursuits. For recent perspectives on the role of assessment in behavior modification, see the reviews by Kaufler (72) and by Goldfried & Kent (58), as well as the important theoretical article by Mischel (104). And for an assessment book that behaviorists will find congenial and informative, see the volume edited by Costello, entitled Symptoms of Psychopathology (21). This handbook includes chapters on the assessment and treatment of such “symptoms” as depression, anxiety, phobias, obsessions and compulsions, tics, stuttering, thumb sucking, enuresis, hyperactivity, sleeplessness, alcoholism, drug addiction, academic underachievement, exhibitionism, homosexuality, asthma, headaches, and other horrors—only a smattering of the ailments of any graduate student.

Indeed, students can often convince themselves that their brains are turning mushy. If this is your own hypothesis, it might be wise to consult one of the recent volumes on the assessment of brain damage, either Goodglass & Kaplan’s Assessment of Aphasia and Related Disorders (60), a test manual for the Boston Diagnostic Aphasia Test, or Russell, Neuringer & Goldstein’s Assessment of Brain Damage: A Neuropsychological Key Approach (130), essentially a test manual for the Halstead-Reitan Impairment Index. Three related volumes are focused primarily on children: (a) a report of a superb empirical study, Deutsch & Schumer’s Brain-Damaged Children: A Modality-Oriented Exploration of Performance (29); (b)
another monograph reporting a study of cognitive perceptual motor dysfunction (Rubin et al 129); and finally (c) Wender's new textbook, *Minimal Brain Dysfunction in Children* (152).

Disturbed children tend to maintain a low profile in the eyes of the community at large, at least until one of them begins to steal things or to assault others, when he becomes a part of that national epidemic called "crime in the streets." Among the many publications which relate at least tangentially to the prediction of criminal behaviors are the theoretically exquisite volume by Schachter, *Emotion, Obesity, and Crime* (132); a revised edition of Eysenck's *Crime and Personality* (41); a monograph on underachievement by Kipnis called *Character Structure and Impulsiveness* (74); and two new books by the high priests of this field, Sheldon and Eleanor Glueck (49, 50). One of their volumes, *Toward a Typology of Juvenile Offenders* (49), reports the results of an empirical study in which three types of delinquents were discovered. The other book, an edited volume on the identification of predelinquents (50), should be read by everyone in assessment—at least by everyone who is willing to be exasperated, provoked, and stimulated at one and the same time. Some of the contributors to this short book discuss the empirical evidence on the validity of the Glueck five-factor table (based on interviewer ratings of (a) kind of paternal discipline, (b) type of maternal supervision (c) degree of paternal and (d) maternal affection, and (e) cohesiveness of the family) in predicting—or, unfortunately, most often postdicting—delinquency. The great mass of "evidence" from retrospective studies indicating that the system correctly identifies (I kid you not) 91%, 81%, 91%, 89%, 90%, 90%, 91%, 92%, 100%, 100%, 82%, 81%, 86% of various delinquent samples in the U.S. and 92%, 92%, 85%, 88%, 90%, 80%, 91%, 90%, 98%, and 92% of various delinquent samples in Japan, Puerto Rico, England, France, and West Germany is clearly not scientific evidence at all, since these studies included data from only two cells in the overall fourfold table, always omitting the proportion of nondelinquents who would have been correctly and incorrectly diagnosed. On the other hand, that rising righteous indignation of yours may be wrongeous, if one is to believe the results of a few genuinely predictive studies reported in later chapters of this book; the apparent validity of the Gluecks' prediction scheme is uncanny—in fact, it seems better than reliability theory would suggest is possible. Clearly, there is something funny going on here, and it's about time that somebody takes a highly critical peek at it all.

If delinquency is a phenomenon which is quite predictable, at least by some standards, then suicide is near the other pole of this continuum. Two reviews of the scientific literature on suicide prediction were published during this period, one by Lester (78) and the other by Brown & Sheran (10). Neither is very encouraging. Nor are Bloch & Goodstein (7), who have reviewed the empirical evidence on the links between personality measures and various speech disorders, including stuttering. Their verdict: no links.

*Diagnosis Based on Naturalistic Observation or Life History Data*

A two-volume potpourri of conference papers, entitled *Life History Research in Psychopathology* (126, 127), provides brief reports of some recent longitudinal investigations. The first volume is focused largely on studies of schizophrenia, while
the second includes a wider sampling of problems (e.g. delinquency, alcoholism, narcotic addiction, and suicide). In an important related paper, not included in these compendia, Overall (109) discussed a 2000-patient exploration of the relationship between marital history and psychopathology (the more previous marriages, the less present pathology, regardless of present marital status). Moral: It's not what you're doing, but what you've done, that counts. This conclusion is hard to assimilate with those from a series of investigations (Myers et al 106, Rahe et al 121) which suggest that the number of life changes one experiences is positively related to the incidence of physical and emotional problems. (Ostensibly, having many marriages and divorces means having a lot of life changes.) However, the life-change findings have come mainly from military samples, where the incidence of extreme psychopathology must be rather low, while Overall's conclusion stems exclusively from a study of psychiatric patients, where the incidence of nondeviancy is definitionally near zero. Investigators looking for something fun to do might try combining the methodology of both approaches, using a heterogeneous sample of subjects.

Back in the old days (meaning right now at most traditional psychiatric settings), diagnosis was a single-stage, global, completely intuitive process: patients were interviewed by a clinician, who then assigned the patient to a diagnostic class or type. Nowadays (meaning at some research-oriented settings), these traditional diagnoses are supplemented by others which stem from a multistage process that combines more molecular intuitive judgments with empirical classification procedures. Specifically, the clinician, at the end of his diagnostic interview (142), now may complete one or more standardized rating scales, recording forms, or behavioral checklists. The clinician's responses to these precoded forms are then combined, by clerk or computer, to provide a set of syndrome scores. The profile of scores, in turn, is then used, again by clerk or computer, to assign the patient to a particular diagnostic cluster or type. The classification system (i.e. the entire set of types) may originally have been developed either intuitively or empirically via previous cluster analyses of large samples of score profiles. An example of the former is Spitzer & Endicott's DIAGNO system (140), while examples of the latter include the taxonomies devised by Lorr (82, 83, 87) and by Overall (110, 112).

These attempts to standardize, to fractionate, and thus to objectify the process of clinical decision making have been among the most significant achievements in psychodiagnosis during the past decade. Yet, since most reports of these developments have been published in psychiatric journals, this work may not be as familiar to psychologists as it should be. For this reason, the present chapter will provide more detail about these procedures than about most others, for which extensive reviews are already available (12).

The standardized report forms, which provide the initial data for all of the semiautomatic diagnostic systems, differ from each other in length, in focus, in the degree of structure they impose on the clinical interview from which the clinician's observations and judgments are derived, and—of most importance—in the level of inference required of the clinician (159). At one extreme of this continuum are the most abstract types of depth interpretations found in some psychoanalytic interviews; at the other extreme are the behavior frequency counts favored by investigators with a social-learning viewpoint.
The widely used Brief Psychiatric Rating Scale (BPRS), developed by Overall and Gorham, lies on the abstract side of this continuum (115). A 7-category scale (not present, very mild, mild, moderate, moderately severe, severe, extremely severe) is used for rating 18 symptoms. In addition, the BPRS can be scored for four higher-order factor scores (thinking disturbance, anxious depression, hostile-suspiciousness, and withdrawal-retardation), each the sum of three separate ratings; for two syndrome contrasts (schizophrenic/depressive and coping/resignation); and for a total pathology composite (the sum of all 18 ratings). Recently, Overall has developed the Factor Construct Rating Scale (FCRS), which uses the same type of 7-category scales for rating 17 symptoms. Cluster analyses of FCRS profiles have suggested 5 major patient types, which Overall has labeled depression, thinking disorder, extrapunitive,ness, neuroticism, and agitation-activity (110, 112).

Lorr & Klett's popular Inpatient Multidimensional Psychiatric Scale (IMPS) demands somewhat less abstract judgments than those required for the BPRS or the FCRS. Its 75 items include 45 symptom intensity ratings (each on a 9-category scale), 13 symptom frequency ratings (each on a 5-category scale), and 17 dichotomous items. Newly hospitalized psychotic patients were interviewed and rated independently by two raters using the IMPS; factor analyses of such data from a number of separate samples provide the data base for 10 IMPS syndrome scores. Cluster analyses of IMPS profiles in various samples have led Lorr to conclude that about 60% of psychotics can be categorized into 11 types, 7 of which are the same for men and women (excited, excited-hostile, hostile paranoid, hallucinated paranoid, grandiose paranoid, and anxious-depressed), 2 are unique to men (disoriented and anxious-disorganized) and 2 to women (excited-disorganized and retarded-disorganized) (82, 83, 87).

In contrast to the rather global taxonomic efforts of Overall and of Lorr, a team of investigators at the New York State Psychiatric Institute has been attempting to map the activities and behaviors of psychiatric patients in a far more articulated manner. Led by Spitzer & Endicott, this interdisciplinary group has produced three major rating schedules and a half dozen recording forms during the period of this review. The Psychiatric Status Scale (PSS) was designed to provide multifaceted coverage of psychiatric signs and symptoms (including the use of drugs and alcohol, plus antisocial and illegal activities), as well as to evaluate impairment in diverse life

---

2Conceptual disorganization, hallucinatory behavior, unusual thought content; anxiety, guilt feelings, depressive mood; hostility, suspiciousness, uncooperativeness; emotional withdrawal, motor retardation, blunted affect; somatic concern, tension, mannerisms and posturing, grandiosity, excitement, and disorientation.

3Mental disorganization, mental distortion, affective impoverishment, bizarre motor manifestations, psychomotor retardation, agitation-excitement, extrapunitive,ness, depressive mood, elevated mood, passive dependance, dominant control, denial, intellectualization, over-responsiveness, neuroticism (obsessive, compulsive, or phobic symptoms), and anxiety-tension.

4Excitement, hostile belligerence, paranoid projection, grandiose expansiveness, perceptual distortions, anxious intropunitive,ness, retardation and apathy, disorientation, motor disturbances, and conceptual disorganization.
roles (143). The 21-page PSS booklet includes both a standardized interview guide and an interrelated inventory of 321 precoded items, most of which are brief non-technical descriptions of small samples of verbal behavior (e.g. Query: "What kinds of things do you worry about?" Item 15: "Mentions he worries a lot or that he can't stop worrying"). Factor analyses of item responses from an earlier version led to the development of 17 PSS symptom scales, a denial of illness scale, plus five role impairment scales (for the roles of wage earner, housekeeper, student or trainee, mate, and parent). In addition, as with the BPRS, the symptom scales can be summarized by four higher-order composite scores (subjective distress, behavioral disturbance, impulse control disturbance, and reality testing disturbance), as well as an index of overall role impairment.

A close relative of the PSS is the new Psychiatric Evaluation Form (PEF), an attempt to cover much the same ground using far more global ratings (38). The PEF is housed in a 9-page booklet, which contains 27 rating scales and 2 checklists. Nineteen symptom scales are scored, plus the same 5 types of role impairment as in the PSS. Another new instrument in this series is the Current and Past Psychopathology Scales (CAPPs), which includes 28 items from the PEF, 13 others of some incremental diagnostic significance, and 130 life history items (37). CAPPs scoring keys, again based upon factor analyses of previous protocols, include 8 indices of current pathology, plus 18 life history scales. In addition, the 171 CAPPs responses can be used to provide a computer-based psychiatric diagnosis, via the DIAGNO-II program (140).

PSS, PEF, and CAPPs, which were designed primarily as research instruments, are multipage booklets with separate scoresheets which must be keypunched prior to computer processing. Such instruments are hard to sell to busy practitioners, whose devotion to paperwork may not be so fueled by scientific ambition. Nonetheless, clinicians must be inveigled into cooperation if the diagnostic scientists are to achieve their dream: a large-scale systematic research project in which standardized observational data are collected on all patients admitted to every hospital in a state.

* Depression-anxiety, daily routine or leisure time impairment, social isolation, suicide or self-mutilation, somatic concern; speech disorganization, inappropriate affect or appearance or behavior, agitation-excitement, interview belligerence or negativism, disorientation or memory impairment, retardation or lack of emotion; antisocial impulses or acts, drug abuse, reported overt anger; grandiosity, suspicion-persecution-hallucinations; and alcohol abuse.

* Depression, anxiety, suicide or self-mutilation, somatic concerns; speech disorganization, agitation-excitement, hallucinations, inappropriateness, disorientation or memory loss; belligerence-negativism, denial of illness, grandiosity, suspicion-persecution; retardation or lack of emotion, social isolation, impairment in daily routine or leisure time; narcotics-drugs, antisocial; and alcohol abuse.

* Reality testing, depression-anxiety, impulse control, somatic concern, disorganization, obsessive-guilt-phobic, elation-grandiosity, and overall role impairment.

* Depression-anxiety, impulse control, social and sexual relations, reality testing, dependency, somatic concern, obsessive-compulsive symptoms, anger-excitability, mania, sexual disturbance, memory and orientation, disorganization, organicity, neurotic childhood, phobia, retardation, hysteria, and intellectual performance.
or better yet, in a multistate region. As a compromise between the scientists’ need for data standardization and response completeness and the practitioners’ need for brevity and simplicity, a number of new recording forms have been developed. Among the most impressive of these is an integrated set of six psychiatric record forms developed by the Spitzer & Endicott team (141). These forms, which were developed for a collaborative project involving 7 states and the District of Columbia (the Multi-State Information System for Psychiatric Patients), were designed for recording psychiatric status and history within a computerized record-keeping system. All are 8 1/2 by 11 inch forms which can be automatically processed by an optical reader for transmission directly into a computer and/or for storage on computer punchcards. The Problem Appraisal Scales (PAS), a one-page summary of the patient’s presenting problems upon admission to the hospital, includes 38 5-category rating scales covering physical, intellectual, and social symptoms. The Psychiatric Diagnosis Recording Form (PDRF) is a single-page checklist for recording one, two, or three diagnoses from the “official” list provided, plus the diagnostic stage (admission, provisional, reevaluation, or final). The Mental Status Examination Record (MSER) is a 4-page condensation of the PSS, which includes 156 symptom rating scales and 101 checklist items. A hospital (PER) and a community (PER-C) form of the Periodic Evaluation Record permit a brief evaluation of patient functioning during the past week; each of these single-page forms includes 28 scales and 45 checklist items from the MSER, thus permitting analyses of patient change. Finally, the Psychiatric Anamnestic Record (PAR), a 4-page form with 152 rating scales and 266 checklist items, covers the information typically included in a complete psychiatric case history.

Another enterprising effort to standardize and automate clinical records is being carried out by Sletten, Ulett, and their associates at the State of Missouri’s Division of Mental Health (34, 138, 151). This team has been developing a computer-based network called the Missouri Automated Standard System of Psychiatry, which now serves 10 hospitals and 9 clinics within the state. Virtually everything normally collected from and about each patient is fed into the computer’s hungry maw, mostly via single-page checklists which substitute for the transcribed records of yesteryear. The Missouri Automated Mental Status Examination (136) includes 111 symptom-descriptive items which are scored on two dozen or so factor analytically based scales. These scales plus another couple of dozen demographic variables have been used in stepwise multiple-discriminant analyses to predict the final psychiatric diagnoses for 12 types of patients (139). Studies using large samples and a cross-validation design have suggested the utility of these linear functions for purposes of diagnosis, as well as for predicting length of stay in the hospital, likelihood of running away, and potential for committing suicide and for assaulting others (134).

Other single-page Missouri checklists include: (a) the Missouri Inpatient Behavior Scale, 90 items relating to ward behavior; (b) the Community Adjustment Profile System, 100 items about social functioning obtained from the patient’s relatives or friends, which provides 10 scale scores; (c) a form to record demographic data plus all movements between installations; (d) a checklist for the
emergency room or admissions desk; (e) one to record the findings from a physical examination; (f) a form for the social worker to record information about social functioning; (g) an alcohol questionnaire; (h) a medicaid report; (i) an outpatient contact sheet; (j) a form for recording all medication data; (k) another for recording blood chemistry; (l) yet another for information about any nursing home care; and (m) various forms for business records. Even the Missouri psychometrist is in the act. While the Peabody Picture Vocabulary Test (a standard instrument in the system) is not yet administered by robots, patients sort the 550 MMPI items, each printed on a prepunched IBM card, into two piles (True vs False), which are then fed to the computer for scoring. With such a system it becomes possible to check a clinical hunch with a flick of the programmer. Want to know what sort of patients committed suicide in the state during the past few years? Just ask the computer to search its files, find them all, and list their combined characteristics. Voila (135)!

The Institute of Living in Hartford, Connecticut, is another institution that has attempted to computerize evaluations. Their new instruments include: (a) a 650-item mental status examination which can be completed by the clinician via a television monitor integrated with a modified typewriter keyboard (Donnelly et al 32); (b) a one-page machine-readable form for recording nursing notes, thus permitting an automated sequential analysis of patient change (Stroebel & Glueck 144); and (c) a single-page machine-readable interview schedule that provides delinquency predictions in a more objective manner than has been possible in the past (Glueck 48). All of these efforts, however, are based on the use of prespecified coding categories (e.g. multiple-choice questions). Wouldn't it be nifty, some have suggested, if the clinician or the patient could respond to the computer in their natural language, in narrative essay style? That day hasn't quite arrived, but it may be on the way. Eidsun & Ramsey-Klee (35) at the Reiss-Davis Child Study Center in Los Angeles have been designing a general information processor, called the Psychiatric Case History Event System (PsyCHES), which represents a step in this direction. And Benfari et al (6) at Harvard have been developing another such sophisticated system, called Computer Assigned Symptom Evaluation (CASE).

If the reader hasn't fallen asleep by this point, news of some additional L-data measures may serve to do the trick. A short symptom checklist for the diagnosis of schizophrenia, the New Haven Schizophrenic Index (NHSI), can be scored from hospital records; intuitively derived item weights are used to produce a total score, which purportedly is highly related to clinical diagnosis (3). Also working from patient records, Hattulauma (67) has reported a methodologically sophisticated analysis of psychiatric syndromes and their relation to antecedent behaviors and treatment prognoses. In a similar vein, Martorano & Nathan (90) factor analyzed the 100-item Boston City Hospital Behavioral Checklist and related these factors to those from the IMPS and to traditional diagnoses. Similar types of analyses have been undertaken to compare the data from American versus English patients (Eve-rigg et al 39) and to identify and/or classify depressed patients (Paykel 117, Zung 160). And when a 64-symptom instrument, the Symptom Distress Checklist, was factor analyzed, items in the resulting four factors were found to be highly related to the clusters obtained when 20 clinicians were instructed to group the 64 symp-
OBJECTIVE DIAGNOSTIC TESTS AND MEASURES

Although the vast majority of these investigations have been focused exclusively on adult psychiatric patients, there have been a few studies of children. Among the most intriguing new child measures is the Personality Inventory for Children (PIC), which consists of 600 true-false items to be completed by the mothers of children between the ages of 6 and 16 (Wirt et al. 158). PIC scales, which are still under development, include 11 constructed by the intuitive strategy,9 plus at least 8 constructed by the external (empirical) approach.10 Far more restricted in content is the Louisville Fear Survey for Children, an 81-item inventory that can be completed by parents or teachers on the one hand, and by children or their peers on the other (Miller et al. 103); a factor analysis of parents' ratings of their children suggested three major kinds of fears: physical injury, natural events, and psychic stress. Teachers' ratings of the problem behaviors of elementary school children have been factor analyzed by Miller (102) and by Dielman & Cattell (30, 31). Miller's analyses have led to the development of his School Behavior Check List, which includes 6 subscales (low need for achievement, aggression, anxiety, academic disability, hostile isolation, and extraversion), plus a total disability score. Dielman & Cattell found 5 replicated oblique factors from a 62-item problem behavior checklist (hyperactivity, sluggishness, suspiciousness, withdrawal, and acting out). Modern kids must be turning soft; there were far more ways of getting into trouble back in the good old days.

Diagnosis Based on Questionnaire or Self-Report Data

Life history data normally must be encoded by human observers before being processed for assessment purposes, and we human observers have been known to have less than perfect fidelity for this transmission task. Various biases, context effects, and stereotypes plague us; our channel capacity is severely limited; and we get disrupted by boredom, fatigue, illness, as well as a host of situational and interpersonal distractions. In short, since humans are expensive middlemen, psychometricians often opt for self-report measures, which attempt to deal direct. However, an individual's social stimulus value is a mighty significant aspect of his personality, and individuals may not always see themselves exactly as others see them. So personality assessors pays their money and takes their choice: between L-data, smudged, warped, or otherwise distended by gitches in the transmission system; or Q-data, with its own full share of foibles and frailties. If we are wealthy enough, we can buy both products. And when we can figure out how best to meld them together, we will have achieved a scientific breakthrough of the highest magnitude (57).

9Withdrawal, excitement, reality distortion, aggression, somatic concern, anxiety, social skills, family relations, physical development, intellectual development, and asocial tendencies.

10General maladjustment, delinquency, learning disability, sexual identification, cerebral dysfunction, intelligence, somatization, and defensiveness.
For the nonce, however, let us consider Q-data as a codified form of interpersonal communication used by one person to evaluate, assess, classify or diagnose another. The content domain of the communication, which is specified by the assessor, can include perceptions of observable past events (9), symptoms of emotional disturbance (51, 92, 131), transitory moods (93), or any other beliefs, opinions, or attitudes of interest to the nosiest of snoops. For example, the Minnesota-Briggs History Record is an attempt to measure via questionnaire some of the variables typically obtained from L-data (9). Indeed, the inventory was initially developed for use by relatives of patients hospitalized for psychiatric problems, and only recently has it been converted to a self-report format. The new version, which contains 127 items, provides 7 scales (family disunity, conflict with parents, health awareness, introversion, school or job failure, social misfit, and breakdowns or addictions). While the Minnesota-Briggs was designed to measure enduring life history patterns, a new Profile of Mood States (POMS) was built to assess highly transient and fluctuating affective states (93). The POMS consists of 65 items, mostly state-descriptive adjectives (e.g., "shaky"), each to be rated on a 5-category scale (not at all, a little, moderately, quite a bit, extremely). Factor analyses of a larger set of items led to the development of 6 highly homogeneous scales, measuring the moods of tension, depression, anger, vigor, fatigue, and confusion. For a fascinating methodological footnote on the measurement of moods, see the report by Meddis (95).

Psychodiagnostic interviews, for all their potential utility (142), have one clear liability, namely their extraordinary cost in professional time. The desire for an inexpensive substitute for a diagnostic examination led to the development of the very first personality scales (54) and subsequently the need seems to have continued unabated. Maultsby & Slack (92) have developed a computer-based system for obtaining case histories directly from the patient, the output being a report to the clinician. However, computers themselves are not all that cheap, and they are awkward to truck around on house calls. Clearly, a self-administered questionnaire would be the cheapest way out. D. P. Goldberg (a distant relative of the author, on the O'Hallahan side of the family) has recently attempted to construct another one, this to identify individuals with nonpsychotic psychiatric disorders and to assess the severity of their problems. The research monograph that reports his efforts (51) includes the response proportions from each of three types of subjects (psychiatric hospital inpatients, clinic outpatients, and a normal community sample) to each of 140 symptom-descriptive items, the overwhelming majority of which differentiated significantly among the three samples. The results of a factor analysis of the 93 most differentiating items are also provided, along with data on the characteristics of a revised 59-item scale. Schachter (131) has been trying much the same trick, with the hope of identifying the schizophrenics among those pregnant women who register for prenatal care. And Selzer (133) has developed a 25-item Michigan Alcoholism Screening Test (MAST), which appears to work wonders at detecting those who admit drinking a good deal (105).

Two multiscale psychopathology inventories have appeared of late, one of which has been published in the scientific literature and is thus free to the user, while the other has been published commercially and is indeed quite expensive. However,
unlike fast automobiles and fine wine, in this case the cheaper is also the better. Specifically, Lanyon (77) has given away his \textit{Psychological Screening Inventory} (PSI), a carefully constructed 130-item instrument that provides scores on 5 scales: (a) alienation (similarity to psychiatric patients); (b) social nonconformity (similarity to incarcerated criminals); (c) discomfort or anxiety; (d) expression (extraversion or undercontrol); and (e) test-taking defensiveness. In contrast, one writes to a post office box in New York's Grand Central Station to find out about the new \textit{Experiential World Inventory} (EWI), a 400-item wonder that is supposed to provide scores on 8 scales: (a) sensory perception; (b) time perception; (c) body perception; (d) self-perception; (e) perception of others; (f) ideation; (g) dysphoria; and (h) impulse regulation (36). For those who must see it to believe it, Table 1 lists some EWI items, along with a few concocted by the famous humorist, Art Buchwald. Actually, Buchwald's items are included simply for comic relief; every \textit{Annual Review} chapter needs some comic relief. But the tabled EWI items are not all that unrepresentative of the entire set, a fact that makes the EWI manual a gold mine to item buffs. For in that manual are all 400 item endorsement proportions, separately computed within various samples of college students, neurotics, schizophrenics, alcoholics, and prisoners. As an example, take that subtle item for vampire detection (312: “I would like to drink blood”). Roughly 1 or 2 percent of the individuals in normal samples endorse that item, while between 5 and 10 percent of those in schizophrenic samples do so. Happily, however, roughly the same endorsement proportions apply to EWI item 48 (“I cannot fully open my mouth”). Lest you get complacent, however, I must pass on the dire findings for item 102 (“I sometimes feel an urge to bite somebody”): 25% of male college students endorse that one, while only 1% of the coeds do so. Add one more problem for Women's Lib.

\textbf{AND THEN THERE'S THE MMPI} As alert readers may have remembered, there is always a lot going on in this arena, far too much to write about here. However, a few major developments probably should be mentioned, if only briefly. Among the most significant recent MMPI studies are three by Overall and his associates. Two of these were on alcohol abuse (116, 153), and the third was a match race between the MMPI and the 16 PF, comparing their ability at differentiating among 13 different types of psychiatric diagnoses; the winner was the MMPI, by many lengths (157). In another excellent study, Lorr & Gilberstadt (84) joined forces to assess the relationships between Lorr's \textit{IMPS} rating types (L-data) and Gilberstadt's MMPI profile types (Q-data); the verdict: very little overlap. Nonetheless, the development of new profile types continues (63), as does research on the characteristics of persons in each type (66, 80, 119). Hopefully these findings will be incorporated into the computer programs that are used to provide those automated interpretive reports, but don't hold your breath. O'Dell (108) has shown that, at least for the products of one of these services, individuals are unable to distinguish between the accuracy of a flattering universal report and one prepared specifically from their MMPI responses. Three overviews of this business, by Fowler (47), by Kleinmuntz (75), and Manning (89) are now available, thus sparing me the pleasure.
Table 1 Would the real set of items please step forward?

<table>
<thead>
<tr>
<th>Set 1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Set 2&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>The sight of blood no longer excites me.</td>
<td>Animals often try to fool me.</td>
</tr>
<tr>
<td>It makes me furious to see an innocent man escape the chair.</td>
<td>I feel like killing untidy people.</td>
</tr>
<tr>
<td>When I was a child, I was an imaginary playmate.</td>
<td>I sometimes taste sound.</td>
</tr>
<tr>
<td>I am bored by thoughts of death.</td>
<td>Someone is making copies of me.</td>
</tr>
<tr>
<td>I become homicidal when people try to reason with me.</td>
<td>I sometimes think other people's thoughts.</td>
</tr>
<tr>
<td>I don't like it when somebody is rotten.</td>
<td>I am afraid somebody may cut off my nose.</td>
</tr>
<tr>
<td>Most of the time I go to sleep without saying goodbye.</td>
<td>Old men are indecent.</td>
</tr>
<tr>
<td>Frantic screams make me nervous.</td>
<td>I am someone else.</td>
</tr>
<tr>
<td></td>
<td>I would like to drink blood.</td>
</tr>
<tr>
<td></td>
<td>I feel that my ideas may turn into insects.</td>
</tr>
</tbody>
</table>

<sup>a</sup>Written by Art Buchwald, the noted humorist.  
<sup>b</sup>Included in the *Experiential World Inventory.*

Many users have complained that the MMPI is too long. Since a good many items, especially those towards the end of the booklet, are not scored on the standard clinical scales, there have been numerous attempts to expunge them. Various expurgated versions of the MMPI have been devised, the most popular being editions of 399, 373, and 366 items. A recent trend is to whittle down this item pool even more, to the first 168 items, or to a selected 86 (the Midi-Mult), or even to 71 (the Mini-Mult) (65). Since the direction of this movement is evident, the future should see the MMPI reduced to bare bones, namely a few single-item scales. For detecting latent obesity, one needs only the lone item: “I can't believe I ate the w-h-o-l-e thing.” The masculinity-femininity scale can be replaced by one true-false item: “I am a male” (54). And the neurotic scales can all coalesce into the single item: “I am a jerk.”

Whatever one thinks of the present status of the MMPI, virtually no one will deny its extraordinary historical influence upon assessment research and upon psychodiagnostic practice. Eventually, however, old inventories, like old soldiers, must slowly fade away. Hopefully the next decade will bring the MMPI's successor, perhaps the new *Differential Personality Inventory* (15, 147) if it can be lured out of hiding. In the interim, since the MMPI will continue to be used and used heavily, it should be employed as sagely as possible. At present it is used anachronistically; the original clinical scales (each developed empirically to differentiate a particular small sample of psychiatric patients from a larger nonpsychiatric sample) are now employed as the structural elements in fairly elaborate quasiconfigural profile
typologies, and an individual's responses are interpreted in reference to the attributes that tend to characterize others with similar score profiles. However, as Norman (107) has argued so articulately:

Whatever one thinks of the desirability of the original construction of the basic clinical scales (granting the purposes for which they were intended), it is abundantly clear that they are about as inappropriate and maladapted a set as one could imagine for their current uses in profile analysis and interpretation and typal class definition [p. 64].

Specifically, the original clinical scales have at least three major liabilities for the task for which they are now being employed: (a) each scale is highly heterogeneous in manifest content, and consequently it is next to impossible to attribute any unambiguous and/or content-coherent message to a particular scale score; (b) the set of scales is both substantively and structurally redundant, a problem that is severely exacerbated by items keyed on two or more scales; and (c) a sizable amount of potentially significant information from the initial item pool is not even available, either because it stems from the many items that are never scored or because sets of negatively correlated items are included in the same scale, thus effectively eliminating their effects. While these psychometric facts of life were not obvious in the late 1930s when the MMPI was being constructed, they are today. Consequently, until the MMPI item pool is replaced, it is time to move from the old clinical scales to a new set of content-coherent and structurally independent ones.

Over the years there have been numerous attempts to factor various sets of MMPI items. In contrast to similar analyses of MMPI scale scores which typically extract around three factors, the various analyses at the item level have produced little agreement about the number of factors in the MMPI pool; the estimates have ranged from around three, when the late Robert Tryon tried his hand at this game (148), to some unspecified but far larger number in the rototplots of Raymond B. Cattell (27). Barker, Fowler & Peterson (4) extracted nine factors, but then they only used 373 of the MMPI's 550 items. In any case, given a few dozen extra pages, I would argue the heresy that factor analysis of items is as risky an enterprise as is empirical item analysis using criterion groups. What is needed for the construction of new MMPI scales is a multistrategy approach that is maximally sensitive to item content.

Wiggins (154) has used precisely such tactics in the development of his 13 new MMPI content scales. Always with a sharp eye on manifest content, while simultaneously keeping his sights aligned by means of empirical homogeneity analyses, Wiggins has fashioned a set of MMPI scales that are far better suited for use in profile typologies than are the original clinical scales. Unfortunately, since the content scales haven't been around as long, there are far fewer data available about them (118, 146, 156). As a result, the practitioner should be encouraged to employ both the clinical and the content scales, just as with the Strong Vocational Interest Blank he can now employ the statistically homogeneous and content-coherent Basic Interest Scales along with the classic occupational ones (8, 14). And, to past MMPI investigators, an urgent request: rescore those old answer sheets on Wiggins' content scales, and run that vintage study once again. Two publications for the price of one! If anyone should object, tell them you obtained your authorization on these very pages.
Diagnosis Based on Test Data

Inventories are often loosely referred to as personality tests, but they aren’t, and thus they shouldn’t be. One cannot be too sticky about this, however, since over the years the term “objective test” has been defined in extraordinarily diverse ways. As this chapter is not the place to resolve the arcane semantic issues involved, tests will be defined, quite dogmatically, by right of adverse possession if you will (or if you won’t, by right of eminent domain): an objective test is one in which the examinee, without gross cheating, can alter his score in one direction at most; this definition is roughly equivalent to that proposed long ago by Cronbach for a “maximum performance” task. By this definition, most standardized measures of physiological activity and physical proficiency, as well as most aptitude and achievement examinations, qualify as objective tests.

Within clinical psychology these instruments have rarely been used for differential diagnosis, with the notable exception of those used for the detection of brain damage. The latter measures really should be reviewed by someone whose knowledge of neurophysiology is a light year or two beyond that of your reporter; consequently, the wary are hereby advised to skip immediately to the next paragraph. For you unwary, however, five excellent studies of the Halstead Neuropsychological Test Battery have recently appeared (59, 76, 79, 123, 124). Overall & Gorham (111) used the WAIS and a group-administered and computer-scored version of the Holzman Inkblot Test to check the hypothesis that performance differences between brain-damaged patients and those with no such damage are similar to the differences that normally occur as a function of aging; for the two tests employed, that hypothesis proved dead wrong. In another report on brain damage in old age, Kendrick (73) discussed his Digit Copying and Synonym Learning tests. Finally, Lewinsohn and his students (81) have tested patients’ short-term memory via visual, auditory, and kinesthetic modalities, showing that the test performance of those with anterior damage is worse than those with posterior. Moral: Fall backwards.

Welcome back, wary. As you probably know, the outstanding, and longstanding, advocate of the use of objective tests in clinical diagnosis is Raymond B. Cattell. Over the years, Cattell has been collating existing tests, devising new ones, and factor analyzing various subsets of his resulting collection. He has published two major volumes dealing with these efforts, as well as two versions of his Objective-Analytic Personality Test Battery. For so energetic a sales campaign, customer response has been sluggish. The problem has been to convince clinicians that objective tests are more valid predictors of significant clinical behaviors than are other types of measures commonly available. Fabian & Comrey (46) made such a comparison and reached the opposite conclusion. Most of Cattell’s efforts, however, always go into yet another factor analysis (17); this ploy, while unlikely to convert the heathen, is certain to stir up someone who enjoys the same bead game (Howarth 69). Finally, however, Cattell has published a report on the relationships between objective tests and clinical diagnoses (19). It should have been a corkscrew. Instead, it must go down as one of the most amorphous and prolix of a long line of publications not always noted for their crispness. The tragedy is that there probably is something important going on here, since Lorr and Hamlin (64, 85, 86) carried out reasonably similar types of analyses and found gold.
FOR THOSE ASSESSMENT SCIENTISTS WHO HAVE BEEN HANGING IN HERE ALL THIS LONG . . .

As usual, there is both good news and bad news. First let me acknowledge the bad news, which you already discovered for yourself before reading word one of this chapter: your best work (your entire production?) has been overlooked in the author's pigheaded scramble to conform to page limitations. The good news: you get another chance 3 years from now. In the interim, what are the scientific problems on which we should all be working? The author has attempted to answer this question, for the broader discipline of which this is a part, in a paper published only a short while ago (57). In capsule summary, these problems include: (a) discovering the most important individual differences from the host observable; (b) measuring those attributes as precisely as possible; and (c) finding optimal methods for employing these measures in both basic and applied endeavors. No wishy-washy character, the author stands 100% behind his previous choice. In the space remaining, let us look at some recent research on the last two of these three problems.

Measurement normally begins with the presentation of a stimulus (e.g. an item or a question) to a subject, in order to elicit a precodified (e.g. multiple choice), or potentially codifiable, response. Such a response is rarely considered a measure in and of itself; a lone response may be both unreliable (in that future responses to the exact same stimulus would not all be identical) and unrepresentative of responses to closely related stimuli (generalization from single instances, like generalizations from single subjects, can be hazardous to your health). Consequently, the responses to a set of stimuli are typically accumulated and then averaged in some manner before reliable measurement is thought to have occurred. The basic problem in scale construction is selecting the particular set of stimuli, the accumulated responses to which are used to produce a score. There are three general approaches to this task: callously disregarding older terminology, I refer to these as the Intuitive, Internal, and External strategies of scale construction (56). Many test constructors have used a combination of two strategies—typically the Intuitive and the Internal (70) or more rarely the Intuitive and the External (62)—and Meehl (98) now advocates a combination of all three. Jackson (70) has provocatively challenged the alleged superiority of the External over the Intuitive strategies, and Ashton & Goldberg (2) have risen to his bait. For two other such empirical comparisons, see Fabian & Comrey (46) and Goldberg (56).

After measures have been constructed, they really ought to be used. For most purposes more than one measure will be used at the same time, so that the measures—like the items before them—typically must be combined in some fashion. Long ago two broad integration possibilities were distinguished: actuarial (or statistical) and subjective (or "clinical"). When they are pitted against each other repeatedly, as in a cockfight, neither manages to kill off the other in a clean enough manner so that all of the betting money changes hands. Most psychologists have come to agree with Meehl (96), who long ago set up the gaming regulations. On the other hand, Holt (68) has always objected to those very rules, and thus he sporadically confesses his refusal to pay up. The major unresolved problems today (other than persuading Holt to do some research instead of exhorting others to do it for him) include, on
the actuarial side, a specification of the sort of prediction functions that are most appropriate for today’s measures, and on the subjective side, a specification of the optimal meld of empirical data with informed intuition.

The first problem has been formulated as a comparison between linear and configural prediction models (22, 52, 55, 98). As the arch-linearist on these battle lines at present, the author hereby asserts his belief that virtually all psychological relationships are going to turn out to be nonlinear and configural in nature. When measures of individual differences become precise enough, then the various nonlinear prediction models will become useful. At the moment, however, our measures are still too gross to expect nonlinear models to outperform linear ones, given the clear statistical problems involved in the former’s selection and use. These problems are particularly poignant at the moment, since so much effort is presently going into the development of complex profile typologies and into the search for trait-by-treatment interactions; the latter include not only the traditional patient-by-drug interactions, but also some patient-by-therapist ones (20, 122).

And finally there’s human intuition—much exalted, much maligned, often grossly misused, but typically all that’s around. Most sorts of clinical decisions are made subjectively if for no other reason than that at this primitive state of our science the appropriate actuarial data have not yet been analyzed. Ostensibly, practitioners who are called upon to make such decisions on a daily basis come to learn a bit about how to make them. Nonetheless, professionals are still only human and, as noted before, humans have their days: good and bad, up and down, topsy and turvy. A few years ago I pondered whether the practitioner’s inherent unreliability might not be separated from his—hopefully somewhat valid—judgmental policies (53). Since a simple linear regression model can often be constructed to capture critical aspects of many judgmental strategies, what would happen if one substituted one of the judgmental models for the expert himself? An empirical study of 29 clinical psychologists attempting to differentiate the MMPI profiles of psychotic from neurotic patients suggested that models tended to function at least as well as the humans that were modeled (53).

Happily, this finding has now been replicated in a number of subsequent studies, and various investigators seem to have arrived independently at much the same conclusion, albeit by somewhat different routes. For example, investigative teams in Texas (113, 114) and in Missouri (137) have begun to model the decisions of psychiatrists selecting the most appropriate drug for different types of patients. Since diagnosis has already been automated in some settings, this consequent automation of treatment decision making could serve to eliminate the clinician altogether. If you find yourself being tickled by this fantasy, just remember: you could be the very next to go . . .
OBJECTIVE DIAGNOSTIC TESTS AND MEASURES

54. Goldberg, L. R. 1971. A historical survey of personality scales and inventories. See Ref. 94, 293–336
59. Goldstein, G., Shelly, C. H. 1972. Statistical and normative studies of the Halstead Neuropsychological Test Bat-
tory relevant to a neuropsychiatric hospital setting. Percept. Mot. Skills 34:603–20
82. Lorr, M. 1970. A typological conception of the behavior disorders. See Ref. 88, 101–16
90. Martorano, R. D., Nathan, P. E. 1972. Syndromes of psychosis and nonpsy-
chosis: Factor analysis of a systems analysis. J. Abnorm. Psychol. 80:1-10
158. Wirt, R. D., Hampton, A. C., Seat, P. D. 1972. The psychometric prediction of delinquency. See Ref. 127, 66–76