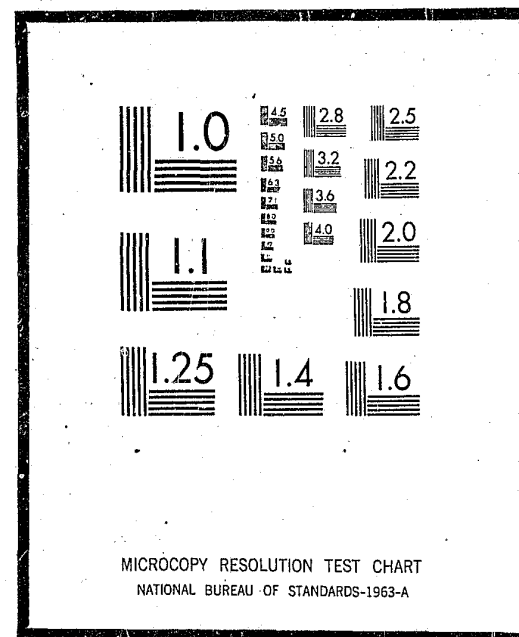


NCJRS

This microfiche was produced from documents received for inclusion in the NCJRS data base. Since NCJRS cannot exercise control over the physical condition of the documents submitted, the individual frame quality will vary. The resolution chart on this frame may be used to evaluate the document quality.



Microfilming procedures used to create this fiche comply with the standards set forth in 41CFR 101-11.504

Points of view or opinions stated in this document are those of the author(s) and do not represent the official position or policies of the U.S. Department of Justice.

U.S. DEPARTMENT OF JUSTICE
LAW ENFORCEMENT ASSISTANCE ADMINISTRATION
NATIONAL CRIMINAL JUSTICE REFERENCE SERVICE
WASHINGTON, D.C. 20531

Date filmed

10/1/75

FINAL REPORT

DOCUMENTATION OF TESTS USED IN OFFENDER CLASSIFICATION

Prepared by:

IIT RESEARCH INSTITUTE
Criminal Justice Science
and Technology Center
10 West 35th Street
Chicago, Illinois 60616

Prepared for:

U. S. DEPARTMENT OF JUSTICE
Law Enforcement Assistance Administration
National Institute of Law Enforcement
and Criminal Justice

JANUARY, 1975

DOCUMENTATION OF TESTS
USED IN
OFFENDER CLASSIFICATION

PROJECT STAFF

Jack P. Kornfeld, Director
Thomas G. Eynon, Senior Scientist
William Bates, Consultant
George S. Speer, Consultant

This project was supported by Contract No.
J-LEAA-023-74, awarded by the Law Enforcement
Assistance Administration, U. S. Department of
Justice, under the Omnibus Crime Control and Safe
Streets Act of 1968, as amended. Points of view
or opinions stated in this document are those of
the author(s) and do not necessarily represent the
official position or policies of the U. S. Depart-
ment of Justice.

JANUARY, 1975
IIT RESEARCH INSTITUTE
Criminal Justice Science & Technology Center
(FINAL REPORT)

ABSTRACT

The monograph is a survey of tests and procedures currently used in correctional practice for classification. The tests are described and their purpose, administration and scoring explained.

The tests are classified as individual or group, and as character and personality; intellectual functioning and organicity; acquired skills; vocational aptitude; interests and values and social adjustment. After a brief discussion of validity and reliability, the monograph covers the following specific tests for correctional interests:

A. Individual Personality Assessment:

- | | |
|-------------------|----------------|
| 1. Bender-Gestalt | 4. TAT and CAT |
| 2. DAP | 5. Rorschach |
| 3. HTP | 6. I-Level |

B. Group Personality Assessment:

- | | |
|----------|---|
| 1. MMPI | 4. CPI |
| 2. 16 PF | 5. Jesness Inventory
Behavior Check List |
| 3. HSPQ | 6. TSCS |

C. Intellectual Measures:

1. WAIS and WISC

D. Measures of Educational Level:

- | | |
|---------------------------|---------|
| 1. Otis-Lennon | 4. WRAT |
| 2. Stanford Achievement | 5. PPVT |
| 3. California Achievement | 6. GATB |
| 7. NATB | |

E. Assessment of Social Adjustment:

- | | |
|-----------------|------------------------|
| 1. Quay Battery | 2. EDS, MRB, WAR, LESS |
|-----------------|------------------------|

The monograph concludes with the following recommendations:

- Base Expectancy Tables using the California material.
- The Reading Subtest of the CAT -- if the reading level is below 9th grade. A special battery will be used of the WRAT and the WAIS if retardate is suspected.
- Intelligence -- if reading level is appropriate, OLMAT; if performance is poor, the Revised Beta is given.
- Personality and Character -- MMPI, 16PF on sample basis.
- Occupational -- GATB; if reading level is low, NATB.

The monograph ends with a short discussion of the problems involved in classifying minors and recommending a continuing research program.

TABLE OF CONTENTS

	<u>Page</u>
I. INTRODUCTION	1
A. Documenting the Tests	6
B. Standards for Offender Testing	6
II. CLASSIFICATION OF TESTS, SCALES AND INVENTORIES	10
III. RELIABILITY AND VALIDITY	25
IV. INDIVIDUAL PERSONALITY ASSESSMENT	37
V. GROUP PERSONALITY ASSESSMENT	49
VI. INTELLECTUAL MEASURES	88
VII. MEASURES OF EDUCATIONAL LEVEL	94
A. Career Planning Measures	97
VIII. EXPRESSION OF INTEREST AND VALUES	101
IX. ASSESSMENT OF SOCIAL ADJUSTMENT	104
X. A SUGGESTED PROGRAM OF TEST AND ASSESSMENT FOR ADULT OFFENDERS	112
A. Base Expectancy Rates Determined By Prior Career	114
B. Education	115
C. Intelligence	116
D. Personality and Character	117
E. Occupational	118
F. Some Recommendations on the Classification of Minors	119
G. Additional Comments	122
APPENDIX--PRIMARY REFERENCES FOR TESTS	124

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1	EXAMPLES OF TESTS, INVENTORIES, OR SCALES CLASSIFIED BY PURPOSE AND KIND OF STIMULUS OR METHOD OF ADMINISTRATION	17
2	STUDIES USING THE BENDER-GESTALT TEST	44
3	STUDIES USING THE TAT	45
4	STUDIES USING THE RORSCHACH TEST	46
5	STUDIES USING THE MMPI	69
6	STUDIES USING THE 16 PF	76
7	STUDIES USING THE HSPQ	77
8	STUDIES USING THE CPI	78
9	STUDIES USING THE WAIS	91
10	STUDIES USING THE WISC	92
11	STUDIES USING THE CAT	99

PREFACE

Attempts to classify those who offend against society, and thus to explain this deviate behavior have extended over a long history, and have ranged over a wide variety of methodologies and techniques. Possession by demons, the effect of stars and planets, or the influence of the gods were early explanations. As man's understanding of his universe developed, and science emerged from superstition, other more rational explanations and classificatory systems developed. Lombroso's physical stigmata and the modern sociologists' search for factors in environmental deprivation are merely two examples of widely divergent attempts to answer the same question.

We have not attempted to develop an explanation for the causes of crime, nor even to formulate a system for classifying the criminal. Rather, we have attempted to survey the nature of the psychological tests used by others in studying and classifying the convicted offender. Three of the authors have been professionally employed in correctional institutions over a period of many years. From their experience, and their knowledge of the practice of others in similar institutions, they have developed the list of tests reviewed here.

The discussion is not intended to be encyclopedic, and it is not primarily intended for the scientifically trained professional. There are literally thousands of tests available for use

for almost any conceivable study of some aspect of criminal behavior. The variety of test instruments available and used by someone in some setting, is endless.

What we have attempted to do is to discuss tests used in correctional institutions that meet these criteria:

1. Are widely used by significant numbers of practical persons for real correctional purposes.
2. Are relatively easy to administer, score and interpret.
3. Can become a significant part of a planned national system of evaluation, as an essential segment of a model offender classification system.

Many others will feel that some test not in our list should have been added, or should have been listed in place of some test we have included. Some will, perhaps, feel that tests we have included should have been omitted. We can only reply that this is our list, based on our experience, observation, and practice.

We have attempted to describe each test in such a way that even a person unfamiliar with it can have a sound basis for deciding whether that test can be useful to him. For this reason we have described the test, its purpose, its administration, scoring, interpretation, and the results obtained by others who have used the test. An appendix includes such important details as publisher and price.

In reporting the experiences of others with a specific test, we have consciously been selective. For some tests there are thousands of published references from which to draw. For others there are only a few. We have selected studies which we believe are typical and significant. The serious student of any one test or tests may be referred to the literature, or such publications as Buros' Mental Measurement Yearbook to pursue his interest further, and in detail.

Similarly, we have included a brief and simple discussion of reliability and validity. This section is not intended to include a comprehensive treatment of these subjects, but to emphasize the necessity of considering these elements in selecting a test, or in evaluating reported results. The sophisticated, professional user of tests is aware of many details we have omitted; the unsophisticated user of tests needs to be made aware of things to look for, and, if he needs further information, where to look. All tests are not equally effective, and some yardsticks should be provided for the inexperienced, or untrained.

A monograph such as this involves many judgements on the part of its authors. Judgements imply responsibility, which we gladly accept. However, though we accept the responsibility for our judgements, we must consider also the opinions of our colleagues more especially when their judgements do not coincide with ours.

ACKNOWLEDGMENTS

We acknowledge with gratitude the help and stimulation afforded us by those who have critically read our first draft of this manuscript. Many of these have been authors of tests discussed here, some have been authors of tests we have chosen not to discuss. In each case their reviews have been stimulating and enlightening. In some instances we have corrected our conclusions. In others, although considering these comments, our original conclusions have remained unchanged. In all cases, we extend our deepest thanks and appreciation for the efforts and responses of our colleagues.

These colleagues have read and thoughtfully evaluated the first draft:

Dr. Stanley Brodsky	Dr. Carl F. Jesness
Dr. M. DeVine	Dr. Roger T. Lennon
Dr. William H. Fitts	Dr. Ted B. Palmer
Dr. Harrison Gough	Dr. Herbert C. Quay
Dr. Starke Hathaway	Dr. Marguerite Warren

SUMMARY

Classification is a set of procedures which extend back into man's history and been done on a great variety of bases ranging from possession by demons to throwbacks to under-evolved ancestors. This current monograph attempts to discuss those instruments which are currently used in correctional institutions and meet the following criteria:

1. are widely used by significant numbers of practical persons for real correctional purposes.
2. are relatively easy to administer, score and interpret.
3. can become a significant part of a planned national system of evaluation as an essential segment of a model offender classification system.

Each test has been briefly described, and its purpose, administration, scoring and interpretation explained. An appendix summarizes the tests covered, and such items as publisher and cost. This monograph has only presented those studies and references of relevance to corrections. It has not attempted to provide the extensive detail of such works as Buros' Mental Measurement Yearbook. There is also a brief discussion of reliability and validity.

Chapter I - Introduction. Intelligence testing dates back to the 1904 Binet-Simon test and to Terman's 1916 I.Q.

Testing to uncover emotional and personality problems dates back to the work of Woodward in 1917 and has been developed over time into a large repertory of tests, including the Bernreuter Personality Inventory, very important not only in correctional practice but also in general clinical work prior to World War II. Test development in this area has continued to the present with the Eysenck Personality Inventory of 1964 and the Comrey Personality Scales of 1970.

Vocational interest testing started with the interview and counseling guide of Kelley in 1914. Perhaps the two most frequently used inventories of this sort are the Kuder Preference of 1963 and the Strong Vocational Interest Blank of 1969.

In addition to the above types of scales, we have seen developed measures of introversion-extroversion, masculinity-femininity, personal values and needs. Perhaps the most recent of these is the Environmental Deprivation Scale which is a checklist of "criminal offender's environmental inputs" developed by the Rehabilitation Research Foundation in Montgomery, Alabama.

An illustration of how a researcher moves from questionnaire item response to classification is found in the work of Quay and his associates. They developed their Personal Opinion Study which "discovers" four deviant personality types transformed into behavior categories.

A. Documenting the Tests

Identification and location of psychological tests relevant to classification of offenders required conversations and letters to professional colleagues, the scanning of much literature and such references as Buros' Mental Measurement Yearbook. Documentation of tests has included:

1. primary reference, including bibliographical reference, author, and cost;
2. description of test; and
3. application or where and how the test has been used in offender research.

B. Standards

To be useful and valid, tests must be standardized on populations relevant to the individual who is to be assessed. Many tests in general use, and particularly tests in the offender classification process, lack the broad base and adequate sampling to make their results validly useful.

Chapter II - Classification of Tests, Scales and Inventories. Tests may be classified by purpose, materials, method of administration, and so forth. For the purposes of this monograph, we are classifying tests as to whether they are individual or group, and according to trait measured, i.e., character and personality; intellectual functioning and organicity; acquired skills; vocational aptitude; interests and values; and social adjustment.

The behavior or attitude of the administrator of a test may significantly affect the response of the subject. The value of the interpretation of the results is a direct function of the skill and professional competence of the interpreter. Although little professional skill or training is required for the administration of the objective tests, it is assumed here that the testing practices are competent and the test is appropriate to the setting and individual being tested.

The goal of classification is the gathering of information that will permit his assignment to a group for treatment based on common characteristics. The individual is matched to a group and the group then matched to an appropriate treatment program.

The three functions to be fulfilled by the tests are those of identification, classification and research. Any of the tests may be found to have an application to all three functions, depending upon the problem and the skill of the experimenter.

In the evaluation of the tests, one must consider the competing techniques available for use. Validation in the classification situation is more complex than in the test development laboratory. The trait being measured must be pertinent to the purpose, and the test must measure it in a manner suitable to the immediate goals.

The selection of a test is an administrative decision to be made in each testing program on a cost accounting basis involving these factors:

- A. The benefits to the criminal justice system and/or the community of identifying an individual in the target population.
- B. The cost to the system and community of missing one;
- C. The cost to the individual of false identification;
- D. The cost to the system and community of applying the treatment when it is not justified.
- E. The cost in time, personnel and money involved in the testing program; and
- F. The cost of the treatment applied to the target population.

These questions go far beyond the ordinary requirements of test reliability and validity.

Chapter III - Reliability and Validity. Reliability is an expression of the accuracy with which a test measures whatever it measures. Accuracy refers to consistency and stability of measurement. The two most generally used methods are the odd-even or split-halves method, and the test-retest method. The former is a comparison of performance on the odd-numbered items with performance on the even-numbered items. The latter is a comparison of a first administration of a test with the results of a second administration after the lapse of some time period.

A reliability measure often used when evaluating some subjective data such as ratings or behavioral observations is the use of inter-rater agreement. This is really a measure of the

reliability of the observers rather than the instrument. Detailed instructions, training and practice tend to improve inter-rater agreement.

Validity may be defined as the extent to which the test measures what it says it measures. Congruent validity usually refers to the agreement between the test in question and performance on some other accepted test of the purported factor.

Concurrent validity is obtained by comparing groups with established characteristics in their performance on the test. For example, a test of social attitude may be administered to a group of known delinquents and to a group of outstanding "good citizens". A significant difference in the predicted direction would be accepted as evidence of validity.

Content validity may be established by showing that the test successfully measures certain knowledge, traits, skills or abilities that are shown to be necessary in the performance of some task.

Construct validity rests basically on the theoretical formulations that are tested by the instrument. This means that items in an inventory which are responded to in the way that the theory requires are considered valid.

Predictive validity is the degree to which the test or instrument predicts future behavior, and the accuracy of such prediction.

All of these methods of determining validity are legitimate and acceptable under certain conditions. Only predictive studies

meet the hard test of scientific reality: the understanding, control and prediction of behavior.

We must also consider the base rate among the general population of the behavior we wish to predict. If the existence of the behavior is very high in the population, it might be more economical to overlook any differential classificatory procedure and treat the entire population.

In general, test performance seems to be a function of the answers desired. A tendency toward delinquency is more easily predicted than the type of crime the delinquent will commit. There is a danger that the hard-pressed administrator may unwittingly overtax the predictive power of a test in a specific situation for which it was not designed, especially with tests having research potential but not usable for classification purposes.

Although there are computerized systems for interpreting tests, in most situations human judgement is still necessary in integrating the data base and making a correctional decision.

Chapter IV - Individual Personality Assessment. The Bender-Gestalt is essentially the task of copying nine simple drawings. It is one of the most widely used tests for psychologists working in a correctional situation because it is brief, simple and non-threatening, and of value as a test of organicity. Its predictive validities are low, and test-retest reliability are also low.

The Draw-A-Person is also widely used. Reliability and validity values are known and significant. As an intelligence test, its use is appropriate only with children.

The House-Tree-Person is a process rather than a test. It yields I.Q. scores that are highly variable, and considers intellectual function as one aspect of an interrelated total personality constellation. Overall drawing may be indicative of organicity or severe pathology, but individual evaluation is extremely uncertain. But as a clinical tool in the hands of a skilled, trained and experienced examiner it may yield significant clues to the total personality.

The Thematic Apperception Test (TAT) and the Children's Apperception Test (CAT) are tests which present the individual with vague pictures that may be described with an infinite variety of stories. Examiners have come to realize that this is not a test but rather a method of studying personality.

The Rorschach is another test that is subject to the same criticism and limitation as the TAT. The stimulus here is a series of ten cards, some in black and white, and some in color on which appear inkblots to which the individual responds in an unstructured manner.

In general projective tests of personality have low reliability and validity. They are difficult to administer and interpret, and depend almost entirely on the training, experience and sophistication of the examiner. They have little predictive value.

The I-Level Classification System finds its theoretical basis in a paper by Sullivan, Grant and Grant and further developed by M. Q. Grant (later M. Q. Warren). It has been widely used as a method of classifying offenders, as an aid to differential treatment, and in management and assignment decisions, especially in the California Youth Authority. As originally developed, it describes seven levels, but the work with delinquent offenders has essentially been limited to levels I₂, I₃, and I₄. There are a total of nine subtypes within these three levels.

As originally developed, training for the lengthy, clinical type interviews requires a five-week course, and weeks of practice following the training. The system requires further research on populations larger than California Youth Authority wards to show its general usefulness.

Chapter V - Group Personality Assessment. The Minnesota Multiphasic Personality Inventory (MMPI), a 550 different true-false item instrument, is the foremost in the field of objective clinical assessment. It produces scales in nine separate traits and three validity scales. The MMPI comes in a card sorting form, a booklet form and a computerized version. In addition it has versions in other languages including Spanish.

In 1968 Kincannon offered a short version termed the Minimult with only 71 items. Research has shown this to be useful when caution is exercised due to its limited reliability and information potentials.

Several special correctional scales have been attempted, including one on escape and one on violence, but none of them have shown sufficient predicting power to be useful.

The 16 PF Questionnaire is an objectively scored test for individuals 16 and over. It is easily administered and scored and has over 30 years of research behind it. Studies have shown the test to be superior in distinguishing between subgroups such as serious and non-serious offenders. It does not predict institution adjustment.

The Jr.-Sr. High School Personality Questionnaire (HSPQ) is for the age range 12-17. It is easily scored and covers 14 factor or source traits. The inventory has been challenged as deficient in evidence for validity. It is not widely used in corrections and should be approached with caution.

The California Psychological Inventory (CPI) is a 480 item scale, with 200 items taken from the MMPI. It is a self-administered inventory with separate answer sheet and takes 45 minutes of test time. It gives 18 general scales that produce a profile. Test-retest reliabilities are acceptable, and validity demonstrated by correlation with grades and other external behavior. The socialization scale consistently distinguishes between delinquent and non-delinquent groups.

The Jesness Inventory and Behavior Check List are used as objective means of determining I-levels. The Jesness Inventory provides ten scales together with an Asocial Index used to predict delinquency. Reliability and validity data are known and sub-

stantial. The Asocial Index does not effectively discriminate between delinquents and non-delinquents. The inventory identifies but does not predict delinquency, and was not found to predict recidivism in AWOL soldiers.

The Tennessee Self Concept Scale (TSCS) is a 100 item scale that measures self-concept and defensiveness. The self concept of delinquent girls as measured on this scale shows they are more negative, more uncertain, more variable and more conflicted.

Chapter VI - Intellectual Measures. The Wechsler Adult Intelligence Scale and the Wechsler Intelligence Scale for Children (WISC) have been established over the years as a basic psychological diagnostic instrument. It was standardized in a nationwide sample of 1700 adults including a prorated sample of the non-white population. Reliability and validity material are available and substantial. It is the best single measure of intelligence available. Despite some early evidence to the contrary the research literature indicates it is not possible to predict delinquency or mental illness solely from intelligence tests.

Chapter VII - Measures of Education Level. The Otis Lennon Mental Ability Test has been standardized on a sample chosen to represent the educational system, not the country at large. It reflects the highest standards in construction, norming, reliability and validity. It is widely used in institutional classification. It provides grade levels and subject matter grade level equivalent. It also supplies deviation I.Q.

The Stanford Achievement Test has been published since 1922 and has test batteries for grades 1.5 through 12. It is available in Braille or large type and has forms for every standard high school subject and some less common ones. Norms are based on a sample of 22,699. Testing time is 350 minutes.

The California Achievement Test covers grades 1.5 to 12 also. It reports reading, mathematics, and language scores for grades 9 to 12. The norms for the 1970 Edition are based on 203,684 students from all parts of the country.

The Wide Range Achievement Test is a measure of reading, spelling and arithmetic ordinarily given individually. Reliability and validity are high but the test is intended as an adjunct for clinical evaluation, not for general school use.

The Peabody Picture Vocabulary Test (PPVT) consists of 150 numbered plates each with 4 pictures. It is intended for illiterates who point to the pictures that match the words read aloud by the examiner.

The General Aptitude Test Battery is the best multi-aptitude test available for evaluating career potential. It is however, slanted to blue collar occupations. It is really a measure of current status rather than ability to learn.

The Non-Reading Aptitude Test Battery is intended for illiterates, semi-literates, and those from a cultural background different from the traditional native white American. The test is too new to have independent reports of its value in print, but is apparently adequately standardized.

Chapter IX - Assessment of Social Adjustment. The Quay Battery consists of three scales - The Behavior Problem Checklist (BPC), the Checklist for the Analysis of Life History data (CALH) and the Personal Opinion Study (POS). These produce four deviant personality types:

BC1	Inadequate - Immature
BC2	Neurotic - Disturbed
BC3	Unsocialized Psychopath
BC4	Socialized Subcultural

The reliability is good and validity is excellent. Careful factor analysis has provided good internal consistency. It is probably the best test designed for classification. The only question is the relevance of the categories for treatment.

The Environmental Deprivation Scale is an interview guide covering 16 items, each scored "0" for positive inputs, "1" for negative. Total score is the sum. Validated against offense subsequent to release, as measured by the Law Encounter Severity Scale (LESS), there appears to be a significant discriminating relationship. The lower the EDS, the lower the LESS. There is a tendency for high EDS scores to become higher over time. If EDS holds up as a parole prediction device in future studies, it will become a highly valuable scale to use in conjunction with an experience table.

The Maladaptive Behavior Record (MBR) was designed to accompany the EDS as a measure of response outputs. The final measure

in this battery is the Weekly Activity Record (WAR). This has not shown any significant data on difference of means against five groups whose LESS scores are of graduated severity. In general these scales, the EDS, the MBR, and the WAR have weak validity data and rely on subjective judgement, although their dimensions are promising.

Chapter X - A Suggested Program of Test and Assessment for Adult Offenders. To plan a testing program requires a statement of the goal intended. It might be "To obtain maximum useful information with least effort, least time and least cost". We wish to assess the following areas:

- a. Probability of recidivism (base expectancy)
- b. Educational Skills and Background
- c. Intelligence or learning ability
- d. Personality factors and adjustment
- e. Occupational interest, aptitude and ability

A. Base Expectancy Rates Determined by Prior Career

The California Base Expectancy Tables do an adequate job for adult males. There is no similarly well established tables for women or youthful offenders. Since these experience tables change over time, feedback is necessary to change these tables as the population changes. In addition it is imperative that the research be done to construct similar tables for women and youthful offenders. In the development, "normal" populations must be included so that the incidence of "normal" crime can be found.

B. Education

Since reading skill is required for many other tests, the first test to be given should be the Reading subtest of the California Achievement Test. If the reading level is below ninth grade, a special diagnostic battery will be used.

Diagnostic battery: Wide Range Achievement Test to evaluate basic reading, spelling, writing and arithmetic. The WAIS to evaluate intelligence if retardation is suspected.

C. Intelligence

For persons of adequate reading ability, the OLMAT will be used. If the OLMAT shows inferior performance, especially if language difficulty is suspected, we recommend the Revised Beta.

D. Personality and Character

The Basic test will be the MMPI. If time and budget permit, the 16 PF will be added. If these indicate serious maladjustment, individual assessment, to include the Bender-Gestalt and the HTP should be provided.

E. Occupational

No test or test battery appears to yield as much information for the same investment of time and money as the General Aptitude Test Battery. It should be realized, however, that the GATB is oriented to blue collar occupation. If the individual is deficient in educational skills, the NATB can be substituted.

F. Classification for Minors

There is no battery of tests for minors as well established as those for adult males. Intelligence testing should start with OLMAT. If the person tests low, then the Peabody Picture Vocabulary Test (range 3 years to 18) should be given. For personality assessment, individuals 16 and over of adequate reading levels should be given the MMPI. There is no really adequate substitute for the younger or poor reading youths. It is hoped that the Jesness Inventory and Quay Battery can be given on wide enough samples to establish their relative merits. For learning levels we recommend the CAT.

CHAPTER I

INTRODUCTION

Since 1900, testing in the United States has had its primary development in six major areas: intelligence, academic or scholastic achievement, interests, character and personality, aptitude and values or value judgement.

The roots of modern intelligence testing lie in the mental age concept of the 1904 Binet-Simon test, and Terman's 1916 I.Q. concept. Goddard had translated the early Binet tests into English and used them at the Training School at Vineland, New Jersey. Very early intelligence testing of offenders led Thomas to report in 1915 on results of Binet-Simon tests on 300 prisoners at Portsmouth Naval Prison.⁽¹⁾ Since 52% of the prisoners had mental ages below age 13, it was thought that offenders were mentally defective.

However, this theory was dealt a severe blow when it was noted that the Marine guard selected for above average efficiency, scored at a lower level than the prisoners. This, and subsequent studies with similar results, led to the realization that level of intelligence alone is not the explanation of criminal behavior. Unfortunately, many researchers even today continue to seek a single simplistic explanation.

The symptom list of emotional problems developed by Heymans and Wiersma in 1906 was modified, expanded and refined by Woodworth in his 1917 Personal Data Sheet. This document, with its "new

scoring system" which summarized individual item responses into a single score, became the source for most personality and adjustment instruments developed since. Pressey's X-O tests for Investigating the Emotions (1921), the House Mental Hygiene Inventory (1927), Thurstone's Personality Schedule (1930), Bernreuter's Personality Inventory (1932), the Guilford NPI and the Bell Adjustment Inventory (1934), the Humm-Wadsworth Temperament Schedule and the Washburne Social Adjustment Inventory (1935), the Rundquist-Sletto Minnesota Scale for the Survey of Opinions (1936), the Darley-McNamara Minnesota Personality Scale (1941), the Hathaway-McKinley Minnesota Multiphasic Personality Inventory (1943), the Maslow Security-Insecurity Inventory, the Heston Personal Adjustment Inventory (1949), the Berdie-Layton Minnesota Counseling Inventory (1953), the Eysenck-Maudsley Personality Inventory (Neuroticism) (1959), the Scheier-Cattell Neuroticism Scale Questionnaire and IPAT Anxiety Scale (1961), the Eysenck-Eysenck Personality Inventory (Neuroticism) (1964), and the Comrey Personality Scales (1970) all can be traced through the development of character and personality tests since Woodworth's contribution in 1917.

Vocational interest testing began with Kelley in 1914, when he produced an interview and counseling guide. By 1922, Pittsburgh High School students were responding to Miner's Interest Scales and 1927 saw the development of the Strong Vocational Interest Blank, followed in 1934 by the Kuder Preference Record-Vocational. Thurstone's Vocational Interest Schedule appeared

in 1935, followed three years later by the revised Strong Vocational Interest Blank. The Kuder was revised in 1948 and again in 1956. Other Interest Inventories have appeared (Jenkins, Weingarten, Geist, Curtis, Gordon, Guilford-Zimmerman, Fricke, and Clark) but the Kuder Preference, revised in 1963, and the Strong Vocational Interest Blank, revised in 1969, remain the two most frequently used.

Measurement of introversion-extroversion has been developed considerably since Jung first proposed the concept in 1916. I-E scale construction began in earnest with Heidbreder (1926), Root (1931), Bernreuter's Self-Sufficiency Scale (1933), Allport's Ascendancy-Submission Reaction Study (1939) revision and Eysenck's Maudsley Personality Inventory-Extroversion in 1959. It is the Eysencks' work with their Personality Inventory (Extroversion) in 1964 that has the most relevance for the criminal offender since they believe that offenders have an over-abundance of extroversion.

This suggested the possibility that criminals whose conduct in many ways resembles that of psychopaths might also resemble that group in having high scores on neuroticism and extra-version, i.e. belonging to the choleric quadrant, and recent work by Eysenck has shown that this is indeed so. (2)

Although Terman began his study of masculinity-femininity in 1922, the Terman & Miles Attitude-Interest-Analysis Test published in 1936 was the first extensive exploration of sex and personality. Since that time the M-F dimension has become a popular scale included in most multi-purpose personality tests, including the MMPI, the CPI, the Bell Personal Preference Inventory, the SVIB, Guilford Martin GAMIN, the Kuder PR, and the Comrey CPS.⁽³⁾

The study of personal values began with the translation of Spranger's types of men theory into English in 1928 and was elaborated by the Vernon-Allport Study of Values published in 1931. The Allport-Vernon-Lindzey revision in 1951 was followed by Edwards Personal Preference Schedule in 1953. "Values" were translated into "needs" by Murray and his followers at Harvard culminating in the Adjective Check List of Gough and Heilbrun (1965) and the Personality Research Form of Jackson in 1967. More recently, the Rehabilitation Research Foundation in Montgomery Alabama has translated "needs" into an Environmental Deprivation Scale which is a checklist of "criminal offender's environmental inputs" (needs).⁽⁴⁾

Each of these developments has led to further study of various populations, including delinquents and adult offenders. Many theories and hypotheses have been developed, suggested and tested. Even when the results have been negative, we have moved a step further in our knowledge and understanding. At times, of course, communication has sometimes been hampered by differences in terminology and definition. On the other hand, the question of whether Eysenck's European definition of introversion-extra-version is compatible with the traditional American view has probably led to increased study and research in this personality dimension. So far, however, this research has not been productive in terms of understanding offenders and their behavior.

When testing offenders for classification purposes, we must keep in mind four levels of behavior organization: question responses level, habitual response level, psychological trait level,

and the general type level. Testing offenders, as with any other group of people, gives us information at all four levels of behavior organization. A single test, such as the MMPI provides information about specific single responses to questions and these responses may or may not be characteristic of the individual. If the test is repeated, we discover the habitual response level of the offender if similar responses to the first test are given and the habitual responses can be measured by reliability coefficients. Traits are theoretical constructs based upon observed intercorrelation between a number of different habitual responses. Some psychometricians refer to the trait level of behavior organization as primary factor level. The most general organization of behavior is at the type level where we find correlated traits grouped into types, for example, "the extrovert" or "the neurotic." A further illustration of how a researcher moves from questionnaire item response to classification of criminal types using psychological tests is found in the work of Quay and his associates who factor analyzed the California Psychological Inventory as the basis for the development of their Personal Opinion Study which when administered to offenders "discovers" four deviant personality types (BC1, Inadequate-Immature; BC2, Neurotic-Disturbed; BC3, Unsocialized-Psychopathic; and BC4, Socialized-Subcultural) transformed into Behavior Categories.

A. Documenting the Tests

Identification and location of psychological tests relevant for offender classification requires a great deal of search and retrieval activity. Conversations and letters to professional colleagues uncovered the most frequently used instruments of classification. Finding detailed information about the tests has required scanning Dissertations, Crime and Delinquency, Sociological and Psychological Abstracts, Journals, and reference books, notably Buros Mental Measurements Yearbooks.

Three basic types of documentation were developed:

1. Primary reference; namely, the bibliographic reference, author, publisher, cost, and so on;
2. Description or summary data about the test, such as the number of items, format, reliability, validity, correlates, norms, etc.; and
3. Applications or where the test has been used, sample sizes, sample descriptions, and research studies which have used the test with offenders.

B. Standards for Offender Testing

The American Psychological Association has produced a guideline for the development and use of educational and psychological tests⁽⁵⁾ as this document points out:

An essential principle underlying these standards is that the test user, in selecting, administering, scoring or interpreting a test, should know what he is doing and the probable consequences of his activity. He should, most of all, have a clear idea of why he is testing...

Competence in test use is a combination of knowledge of psychometric principles, knowledge of the problem situation in which the testing is to be done, technical skill, and some wisdom. Although it is not appropriate to tell a test user that he needs particular levels of validity and reliability, it is appropriate to ask him to ascertain that his procedures do result in valid predictions or reliable classifications or otherwise conform to the purpose of his testing. (6)

As psychological tests are used more frequently by criminologists researching offender typology⁽⁷⁾ and as testing is expanded by criminal justice practitioners making program decisions for offenders we must be careful to heed the advice of the American Psychological Association. When testers "get carried away" we find the courts bringing them back to constitutional reality. As offenders realize the implications of testing for differential assignment to treatment resources, we will find correctional testers involved in more court cases. Sussman recently pointed out a problem in the New York Family Court where intelligence tests are used routinely in determining where a child is sent for treatment. If intelligence tests are standardized on samples of children not representative of juvenile court cases, then depressed scores of minority children may represent systematic bias which calls into question the scientific validity of the tester.⁽⁸⁾

The enthusiasm of test users and the very real concerns about the nature of criminal behavior has led to a proliferation of theories and tests. The list of tests that have at one time or another been thought to measure some type of behavior that could lead to a better understanding, treatment and control of the "criminal type" is almost endless. Lack of coordination and organization has resulted in much wasted effort, and small, slow progress.

CHAPTER I
REFERENCES

1. Thomas, G. E., "Symposium on Intelligence Tests. The Value of the Mental Test and its Relation to the Service", Naval Medical Bulletin, Vol. 9, pp. 200-211, 1915.
2. Eysenck, H. and Eysenck, S., Personality Structure and Measurement, Routledge & Kegan Paul, London, 1969, p. 40.
3. McReynold, Paul R., (Ed.) Advances in Psychological Assessment, Vol. 2, Science & Behavior Books, Palo Alto, California, 1971, pp. 324-327.
4. Devine, M. D., et al., The Environmental Deprivation Scale (EDS): The Role of Environmental Factors in the Analysis and Prediction of Criminal Behavior and Recidivism, Rehabilitation Research Foundation, Montgomery, Alabama, 1974.
5. Standards for Educational and Psychological Tests, American Psychological Association, Washington, D. C., 1974.
6. Ibid., p. 6
7. See for example: Ferracuti, F. and Wolfgang, M. E., Psychological Testing of the Subculture of Violence, Rome, Bulzoni, 1973.
8. Sussman, Alan, "Psychological testing and juvenile justice: an invalid judicial function," Criminal Law Bulletin, Vol. 10, No. 2, pp. 117-148, 1974.

CHAPTER II

CLASSIFICATION OF TESTS, SCALES AND INVENTORIES

With our modern tests, whose analysis is frequently highly sophisticated, sometimes even more sophisticated than the development of the test, the individual is presented with a specific stimulus and his behavior is observed and interpreted. The stimulus may be a word, a drawing, an inkblot, an arithmetic problem, or any one of thousands of other possibilities. The response may be highly structured, as in a True-False test; partially structured, as in a sentence completion test, or solving an arithmetic problem; or relatively unstructured, as in drawing pictures of things or persons; and basically unstructured, as in responding to vague pictures or meaningless inkblots.

The large number of tests, developed for many purposes by innumerable test constructors, presents us with a problem in classification. Tests may be classified by purpose, materials, method of administration, function used, timed or untimed, power vs. speed, degree of training required for administration or interpretation, and so on. No scheme will satisfactorily classify all tests. Any classification system will depend primarily on the purpose of the classification. In this study we are primarily concerned with a utilitarian classification scheme, elastic enough to include all the tests with which we are likely to be concerned, yet rigid enough to place tests meaningfully in terms of their function and use. Our classification is presented in the Table 1.

Within each classification cel various subclassifications are possible. Thus, in the study of personality, the Rorschach is normally considered a diagnostic instrument, intended to uncover deep and underlying emotional problems, and to expose hidden personality dynamics in the individual whose manifest behavior is causing problems for himself or others. Warren's I-level classification system is basically a method of describing types of persons who can be expected to behave in predictable ways in specified situations.

There is another distinction, often overlooked: the behavior or attitude of the administrator of the individual or projective tests may significantly affect the response of the subject. The value of the interpretation of results will, of course, be directly affected by the skill and professional competence of the one who makes the interpretation. Although the value of the interpretation of the objective tests will certainly be influenced by the professional competence of the user, the effect is much less likely to be significant. Very little professional skill or training is needed for the administration of these tests.

Of course, it is assumed that testing practices are competent, and that the test is properly selected and appropriate to the setting and individual to be tested. It is neither proper nor appropriate to "assess" a Spanish speaking offender with a test or other device that assumes a high level of competence in the English language. Similarly, it is not proper nor appropriate to assess an aptitude for manual or manipulative jobs using devices that require a high level of verbal fluency and comprehension.

The choice of tests to be used in any situation therefore, depends in part on the practical problem of the professional competence of those available for the study. This may well be the most important factor in deciding on the research design, outweighing other important factors of time, cost, availability of subjects, purpose and so on.

The fact that a test is classified one way or another, here or elsewhere, does not limit its function to that classification. Thus, a test designed to evaluate normal educational progress through the measurement of skills expected to be acquired at various ages or educational levels, may clearly identify social problems, personality malfunction, or intellectual deficits, when interpreted in relation to other test measures, social history items, and so on.

A word of caution. For ease of discussion we have called "tests" all of the various assessment procedures and devices to which we have referred. Technically, many of them are not tests. However, we are following the precedent established in "Standards for Educational and Psychological Tests",⁽¹⁾ and include under this one simple title instruments designed to measure ability, accomplishment, attitudes, interests and so on, whether they are technically tests, inventories, interview aids, biographical data forms, or other kinds of diagnostic devices. As in "Standards", we have used the word "tests" to refer to any kind of measuring or assessment device.

Any test user should be encouraged to become thoroughly familiar with this publication. "Standards" not only provides fundamental principles by which any assessment device can be evaluated, but also some basic cautions regarding the use of tests and interpretations. Anyone who uses test data, interview information, biographical data or other behavioral cues to make a prediction based on his assessment, should be aware of the professional and technical standards presented in this small monograph, through the cooperation of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education.

In practical situations tests, expectancy tables, interviews and the like may often be used in situations that do not permit the maintenance of the highest standards of technical excellence. Awareness of the ways and the extent to which an assessment procedure falls short of perfection is the first step towards improvement. Improved assessment procedures, and improved reports of experience with such procedures, can lead to improved prediction experience.

As psychological tests are used with offenders they serve three broad functions - identification, classification, and research. In identification we are trying to locate high risk individuals for special treatment. Thus a psychotic offender might need psychiatric treatment, or a potential escapee or particularly violence-prone individual might require special security measures. Because the administrative decisions in such cases may involve

costly individual treatments of great consequence to the offender and the community involved, we may demand a greater precision of performance from the tests used.

In classification the goal is not the identification of an offender for individual treatment, but rather the gathering of information about him that will permit his assignment to a group for treatment based upon the common characteristics of the group. Here the treatments vary with the group rather than the individual. The individual is matched to a group and the group then matched to an appropriate treatment program. The demands for precise prediction from our tests may or may not be as great depending on the circumstances.

The research function is a less demanding one. When tests are used to identify high risk groups or individuals, it is not for administrative decisions but for further analysis and study. Let us suppose a custodial population in which the escape rate is 2%. In an attempt to reduce this, we administer a test designed to identify potential escapees in order that they may receive some deterrent treatment. The test is given to the entire population of the institution. Subsequent follow-up reveals that the test separated from the total population, a subgroup or segment in which the escapee rate was 20% much higher than the overall rate of 2%. Obviously the test is picking up something about escapism. This suggests that our test may, with further study, be developed to a point of greater efficiency or at least that it may be used for selecting groups with a high risk loading of potential escapees.

But certainly the test could not be used to select a high risk group for special security precautions since, in order to set special conditions for the 20% (not identified) who were potential escapees, the treatment would have to be applied to the entire group, only 1 in 5 of whom were high risk individuals. The other four were false positives, wrongly identified by the test.

Wenk, Robison, and Smith offer a nice example of the problem from a 1965 study in the California Department of Corrections Research Division.⁽²⁾ The study identified a class of offenders, 14% of whom could be expected to violate parole by an act of violence. This is three times as great as the rate among parolees in general. But even if a perfect intervention technique were available, it would have to be applied to the entire group, including the 86% who did not have this potential for violence. In addition, the high risk group identified had less than 3% of the total parolee group, and hence, contributed only 8% of the total violence, leaving 92% unaccountable.

It is particularly interesting to note that of two statistical consultants on the project, one felt that the project should be classed as a failure because it yielded no practical prediction instrument, and felt that the nature of the problem was such that at present, no adequate solution was possible. The other was more optimistic, pointing out that many interesting leads had been uncovered and that with more research something might be developed. This difference of opinion highlights the difference between the predictive and research approaches.

In filling these three testing functions of identification, classification and research, it seems reasonable to classify the tests used into various categories. Each user will develop categories to suit his own needs and preferences. No category is specific to any one function, and most tests will be found to have an application in all three functions.

Of course, any test may be used as part of a research project. Depending on the purpose of the plan, the nature of the group to be studied, the setting of the study and the treatment program planned, there may well be a dozen reasonably suitable instruments available. The selection of the proper instrument is a reflection of the professional skill of the experimenter.

TABLE 1
 EXAMPLES OF TESTS, INVENTORIES, OR SCALES
 CLASSIFIED BY PURPOSE AND KIND OF STIMULUS OR
 METHOD OF ADMINISTRATION

TRAIT	INDIVIDUAL	GROUP
Character & Personality	TAT, Rorschach I-Level Bender-Gestalt	MMPI CPI, 16PF Jesness
Intellectual Function	WAIS WISC	Otis-Lennon Beta
Organicity	Bender-Gestalt	
Acquired Skills	Wide Range Achievement Test	California Achievement Scholastic Apt. Test
Occupational Aptitude	Purdue Pegboard	GATB Bennett Mechanical Aptitude NATB
Interest & Values	Wide Range Interest Orientation Test	Kuder Preference Inventory
Social Adjustment	Environmental Deprivation Scale	Gough Socialization Check List

1. Character and Personality. These tests involve the identification of high risk individuals in need of individual treatment. Most of them originated in the Mental Health area, where they had a diagnostic function in selecting disturbed persons in need of psychotherapy. This is usually considered an individual treatment with a 1-to-1 ratio of one patient to one therapist. Their use is not limited to psychiatric settings, however, as they might indicate the desirability of individual counseling in the educational or vocational areas. Attempts are even being made to use them for the identification of potential escapees, those prone to unusual violence, etc. Again, the goal is the location of an individual who needs individual treatment.

An example of such a test and such a use is the Minnesota Multiphasic Personality Inventory (MMPI). Scale 8 on this test is the so-called Schizophrenic Scale. If a person scores high on this, say in the 80-90 range, and particularly if accompanied by elevated scores on scales F, 2 and 4, one may suspect the individual is psychotic; although the judgement certainly would be confirmed by clinical inspection of the person before therapy were instituted. There also are MMPI scales to identify potential escapees, recidivists, violence-prone persons, etc., but so far most of these are still at a developmental stage and not very successful in individual prediction, though some appear to have considerable potential usefulness.

These tests give us information about the structure of an individual's personality, his beliefs, his values, his habits, his cognitive and behavioral style. Is he hostile? Is he social or a "loner"? What motivates him? On the interaction between these traits and attitudes and the correctional program may depend the success of his rehabilitation.

Thus, scale 8, when the score is only moderately elevated, may indicate some social alienation rather than outright Schizophrenia. Scale 6 (Paranoia) when extremely elevated might indicate clinical Paranoia but in moderate elevation suggests only a tendency to brooding and suspicion. In general then, the information furnished becomes of significance only when integrated into a total structure with other information, and the net result is not the commitment of the individual to some extreme of treatment.

Personality tests are many and diverse. The California Psychological Inventory (CPI) functions much like the MMPI. The Sixteen Personality Factor Questionnaire (16PF) offers measures of 16 primary source traits such as desurgency-surgency (running from sober and serious to happy-go-lucky and enthusiastic) and artlessness-shrewdness (running from unpretentious but socially clumsy to astute, polished and socially aware). The Tennessee Self Concept Scale measures a person's conception of himself and the satisfaction or dissatisfaction he feels as a result; the Strong Vocational Interest Blank measures vocational leanings; and so on, ad infinitum.

2. Intellectual Function. We are generally inclined to think of tests of intelligence or learning ability in this category, but when the term is broadly applied, we should include also those tests which are intended to be diagnostic of some type of intellectual malfunction. Thus, the Beta examination is intended as a general index of brightness, used with individuals who have some literacy problem, while the Bender Gestalt test, though basically a personality test, may also seek to uncover or detect organic factors associated with some failure in the intelligence system.⁽³⁾

3. Acquired Skills. Often the success of a person's adjustment in any situation will depend upon how bright he is and how much he knows. So we may wish to know the extent of his knowledge and his potential for increasing it. Reading, writing, and arithmetic are basic to the individual's ability to acquire new knowledge in school, or in most skilled and white collar jobs.

4. Occupational Aptitude. These are tests intended to discover or uncover specific aptitudes for occupational training. The underlying rationale is that an individual will be more successful in preparation for and employment in an occupation if he is particularly suited to it, and certainly will be more successful than if he is unsuited to it.

5. Interest and Values. It is difficult to draw hard and fast lines separating interests and values from other aspects of personality, but there does seem to be some basis for distinction at least in emphasis. Most authors in this area emphasize the things or activities in which the individual expresses an interest

or liking. As in the area of occupational aptitude, it is assumed that the individual will do better in an activity in which he has some interest, than in one in which he has no interest, or even a dislike.

6. Social Adjustment. More and more we have come to realize that we cannot predict a person's rehabilitation from his ability and personality alone. Much depends upon the environmental circumstances in which he finds himself. The best predictor of successful rehabilitation seems to be whether the offender has something to return to, whether he can achieve a stable interpersonal setting, as Martin G. Groder so succinctly summarizes it, "a job and a woman".⁽⁴⁾ So we wish measures of his social milieu, the setting in which he grew up, the setting in which he transgressed and the setting to which he will return. The Environmental Deprivation Scale (EDS), a 16 item scale, used in an interview setting to get information about environmental inputs and support is an example, as is the Quay Checklist for the Analysis of Life History (CALH), one of whose factor scales is labeled "Situational" (S).

Any test must also be evaluated in terms of the competing techniques that are available for use. One of these is the use of the Base Expectancy Tables.⁽⁵⁾ These resemble behavior checklists except that they are based on history rather than present behavior. They predict what an individual will do in terms of what he already has done. Cannot a prediction of recidivism be made based upon the number of previous incarcerations an individual

has experienced, and might this be an easier procedure than testing? Is recidivism related to the number of times an offender has used an alias, the age at which he was first arrested, or the length of time he has remained free of arrest? Can a composite table involving a number of such measures predict better than a test can" These are researchable questions to be answered empirically.

In evaluating the use of tests with offenders we must have a clear understanding of the purpose of the testing and a definite statement of the goals toward the attainment of which our tests should make a demonstrable contribution. Validation in the classification situation is a more complex situation than in the test development laboratory. It is not enough to establish that a test measures a certain trait. We must also establish that the trait itself is pertinent to our purposes and that the test measures it in the manner suitable to our immediate goals.

There are certain very practical dimensions on which we measure the success of any predictive test in an applied situation. What is its "hit" rate? How many of the target population are correctly identified or "picked up"? How many are missed or not identified? These are our "false negatives." As important as anything else, perhaps more so, is the number of non-target people falsely identified as being among the target population. We refer to them as "false positives" and must always remember that to them the consequences of such false identification may be as damaging as false identification in a police line-up.

No fixed standards of test performance in terms of desired hits and allowable error in misses and false positives can be set. It is an administrative decision to be made in each testing program on a cost accounting basis involving these factors:

- A) The benefits to the criminal justice system and/or the community system of identifying an individual in the target population;
- B) The cost to the system and community of missing one;
- C) The cost to the individual of false identification;
- D) The cost to the system and community of applying the treatment involved where it is not justified;
- E) The cost in time, personnel and money involved in the testing program; and
- F) The cost of the treatment applied to the target population.

These questions go far beyond the ordinary requirements of test reliability and validity.

CONTINUED

1 OF 4

CHAPTER II
REFERENCES

1. Standards for Educational and Psychological Tests, American Psychological Association, Washington, D.C., 1974.
2. Wenk, E. A., Robison, J. O. and Smith, G. W., "Can violence be predicted?", Crime and Delinquency, 1972, (Oct.), pp. 393-402.
3. Pascall, Gerald R., and Suttell, B. J., The Bender-Gestalt Test, New York: Grune & Stratton, 1951.
4. Holden, Constance, "Butner: Experimental U. S. Prison Holds Promise, Stirs Trepidation," Science, 1974, 185, pp. 423-426. Mentions Martin G. Groder as summarizing a stable interpersonal setting as "a job and a woman".
5. Hoffman, P. B., D. M. Gottfredson, L. T. Wilkins and G. E. Pasela, The Operational Use of an Experience Table, NCCD Research Center, Davis, California, June, 1973.

CHAPTER III

RELIABILITY AND VALIDITY

The concept of test reliability is an expression of the accuracy with which a test measures whatever it measures. It may be considered a statement of the confidence the user may have that he would obtain the same results if the test were to be given again at a later day. Accuracy refers to consistency and stability of measurement.

There are several methods of determining reliability, the most common and the simplest, being to compare performance on odd-numbered items with performance on even-numbered items. This method, usually called "odd-even" or "split-half", obviously yields a measure based on an instrument only half as long as the actual test. As the reliability of a test tends to increase as the length increases, this is corrected by the use of one of several available formulae, such as Spearman-Brown, Kuder-Richardson 20, or Cronbach's.

A better, though more difficult and more demanding, method is the test-retest, in which the test is actually administered a second time, after the lapse of some time period. The availability or lack of availability of subjects tends to be the most limiting factor, though the ease of split-half reliability is undeniably a factor as well.

The effect of familiarity with the test, or practice, may affect reliability to some undetermined effect. This is usually dealt with by administering two different forms of the test, constructed in such a way that they are believed to be equivalent, but different. In such studies, where time and the number of available subjects permits, the experimental population is divided into two groups, one group taking the two forms in AB sequence, and the other group taking them in BA sequence.

Test-retest studies must consider the possible effects of some intervening experience, education, event, or behavior on the subsequent scores. Thus, test-retest studies of educational skills such as reading, may measure the effect of instruction, rather than the reliability of the test. Many studies are, in fact, designed in this way to attempt to measure the influence of such interaction. A personality inventory, for example, may be used to measure attitudes before and after various forms of treatment, or no treatment at all.

A reliability measure often used when evaluating such subjective data as ratings or behavioral observations, is the use of inter-rater agreement in the evaluation, observation, or recording of certain data or judgements. Data of this type are really measures of the reliability of the observers, however, rather than of the instrument. Detailed instructions, training, and practice tend to improve inter-rater agreement.

Validity may be defined as the extent to which the test measures what it says it measures. The history of testing contains

many examples of tests which were found, after their publication, to measure some trait other than the one for which it was intended. Particularly with factor analysis, the naming of the trait becomes an important art.

Validity may be established in a number of different ways to meet existing situations, theoretical considerations, or legal concern.

Congruent validity has long been widely used. This usually refers to the agreement between the test in question, and performance on some other accepted test of the purported factor. Thus, a new intelligence test might seek to establish its validity by demonstrating its agreement with an already established test such as the Binet or WAIS. Obviously, of course, validity actually rests on the assumption of the validity of the first test.

A very similar procedure is concurrent validity. Here, groups with established characteristics are compared in their performance on the test. A test of social attitude may be administered to a group of known delinquents and to a group of outstanding "good citizens". A significant difference in type or degree of response would be accepted as evidence of validity. One difficulty is that it is often possible to obtain such differences for existing groups, but not possible to duplicate the results with other groups. Another difficulty is that sometimes the difference may actually reflect a third factor, which may not be recognized. For example, differences in performance on a test of "mechanical aptitude" between a group of apprentices and a

group of skilled mechanics may actually reflect only experience or training, not a true innate difference in ability.

Content validity may be established by showing that the test successfully measures certain knowledge, traits, skills or abilities that are shown to be necessary in the performance of some task. For example, a test of typing skills may be presumed to be valid if the job to be performed requires typing. The value of this type of validity study rests upon the care and thoroughness of the job analysis performed.

The Equal Employment Opportunity Coordinating Council has set as its standard a minimum reliability correlation estimate of .70 for content valid tests. Although these guidelines are not final at the present time, it is expected that changes will be minor. As many courts have already followed the suggested EEOC Guidelines, it may be expected that this will come to be the generally expected standard in the future.⁽¹⁾

Construct validity rests basically upon theoretical formulations that are tested by the instrument. In practice, for example, this means that items on an inventory which are responded to in the way that the construct requires, are considered to be valid. At times, of course, the construct may follow the test construction. Thus, a factor analysis may produce a group of related items that can be understood only in terms of a newly formulated theoretical proposition. This group of items is then named to conform to the author's formulation.

Predictive validity, which the Equal Employment Opportunities Commission, for example, calls the most desirable, is the degree to which the test or instrument predicts future behavior, and the

accuracy of such a prediction. Because human behavior is so complex, and our tests are still so imperfect, we do not expect complete accuracy. We do, however, ask ourselves the question: "Given this score on this test, what is the probability that in these specified conditions, this person will act in a specific way?" The answer to that question is a statement of predictive validity.

When predictive or criterion-related validity is used, it is generally felt that the correlation coefficient should be large enough so that it has a probability of having occurred by chance not more than one time in twenty, a relationship that is expressed as the .05 level of confidence. The degree of confidence is normally obtained by consulting standard tables which taken into consideration the magnitude of the correlation, the number of variables and the number of subjects or scores.

All of these methods of determining validity are legitimate and acceptable under certain conditions. Only predictive studies meet the hard test of scientific reality, the understanding, control, and prediction of behavior.

The study of the validity of an instrument should begin with a theoretical formulation based on a careful and detailed study of the behavior with which the study is concerned. From this an appropriate instrument of desirable reliability is constructed, with reliability measured both statistically by split-half correlation, but also empirically by test-retest methods. The test is then administered to groups of known characteristics, appropriately

to groups known or judged to possess the characteristic studied and to groups known or judged to be lacking in such characteristics.

If the test survives these steps, it is then administered to new groups, with conditions of treatment or experience systematically varied, and behavioral outcomes predicted in accord with the theoretical framework.

We must also consider the base rate among the general population of the behavior we wish to predict. If the existence of the behavior is very high in the population it might be more economical to overlook any differential classificatory procedure and treat the entire population. Thus, if 98% of the population were potential escapees, it would be more efficient to submit everyone to tight security and not worry about the 2% who would not need it. On the other hand, if only 1% of the population were potential escapees it might not be worth the time and effort to locate them, particularly if the classificatory procedure involved a high false positive rate. Under these circumstances one might accept the 1% base rate as a calculated risk.

Keeping all these things in mind, we can comprehend the complexity of the evaluative task in assessing the use of tests for classificatory purposes in a correctional setting. Tests do some things well, some things not so well, and some things not at all. In general, test performance seems to be a function of the answers desired. A tendency toward delinquency is more easily predicted than is the type of crime the delinquent will commit. The performance of a test for potential delinquency given in a public

school situation for the purpose of initiating preventive counseling measures would probably be more satisfactory than would the performance of the same test were it used in a correctional setting for the purpose of making parole decisions.

The danger is that hard-pressed administrators may unwittingly overtax the predictive power of a test in a specific situation for which it was not designed. This may happen with tests which have some research potential but are not usable for classification purposes.

A number of computerized systems for the interpretation of the MMPI have been developed. These provide computer print-outs ranging in length from one to eight pages or more, and in complexity or depth of analysis from the simple reprint for the client or patient to a very complex and technical report for the use of the professional therapist. The number of scales reported on varies from 27 to nearly 200. The cost, too, shows a wide range: from \$1.50 each for more than five, to \$30 each for the professional report.

The use of the computerized report simplifies the interpretation, standardizes the evaluation, and minimizes the problem of training workers to score and interpret the data. However, we cannot, and should not, expect the computer to do our thinking for us.

All in all it seems obvious that in most situations human judgement is still necessary in integrating the data base and making a correctional decision. An automated, computerized, actuarial decision based upon objective test data may be a fitting

goal, but it lies ahead, a promise for the future. In most of our correctional settings test data must still be supplemented by human judgement.

A simple expectancy table shows the relation between a single variable and some criterion of performance, for example, relationship with male parent and age at first offense. However, few significant decisions are made on the basis of only one predictor, so that a multiple-entry expectancy table becomes more useful. Such tables may be unweighted, simple tabulations of successive bits of data, or may be items weighted as the result of statistically sophisticated techniques. As the average layman would not be expected to understand or utilize such procedures as multiple regression equations, the results of such procedures may be displayed in a somewhat simplified manner as a Base Expectancy Table, or experience table.

The "experience table" as a method of predicting success or failure on parole was a technique apparently first introduced by Hart ⁽²⁾ with subsequent major contributions by Burgess ⁽³⁾, the Gluecks ⁽⁴⁾ and Vold ⁽⁵⁾. In essence, the method involves the systematic search for demographic and life history items which distinguish between violators and non-violators on parole. These researchers found such discriminatory items, and later investigators such as Glaser ⁽⁶⁾ and Manheim and Wilkins ⁽⁷⁾ also report positive experiences, although some may not have been as successful.

At the same time that the field of criminology has been working with "base expectancy tables", to predict parole success, the field of industrial psychology has been studying the effectiveness of "base rates", "biodata" and "biographic history items" in the prediction of job success, credit worthiness, and so on. Unfortunately, there appears to have been little communication or cross-fertilization between the two fields. In some respects the approach and the results appear to be similar, but some differences also appear.

The need for validation and cross-validation is stressed in psychological literature, underlining the frequent chance variations.⁽⁸⁾ Guion has shown a technique for weighing personal history items⁽⁹⁾, but Miner points out that this seldom adds enough to the prediction to make the effort worthwhile.⁽¹⁰⁾ Miner also points out that highly predictive responses are often difficult to explain in any way, which suggests the need for caution in using any theoretical system to explain criminal behavior.

In any prediction a chance level of accuracy must be defined based on the observed frequency of the criterion, and the predictive technique must be contrasted with this chance expectation. Obviously, the less frequent the occurrence of the behavior to be predicted, the more difficult it will be to predict a frequency that exceeds the chance occurrence. However, with recidivism rates of 50 to 60 percent or more, the frequency of parole violation is great enough to provide a statistically adequate base.

The real difficulty lies in the lack of an adequate sample for significant cross-validation studies. There are available an adequate number of life history items which may be considered.⁽¹¹⁾ But obviously, by definition, certain items which can be expected to occur in the history of an offender population will not be found in the histories of a sample drawn from life insurance salesmen! Nevertheless, a true validation, and eventual strengthening of the usefulness of a base expectancy table will require an attempt in depth to contrast as many items as possible with normal and offender population. Cureton provides one mathematical model for establishing the true base rate in such a study.⁽¹²⁾

CHAPTER III

REFERENCES

1. Uniform Guidelines on Employee Selection Procedures, Equal Employment Opportunity Coordinating Council, Commerce Clearinghouse, 1973, pp.3335-3361.
2. Hart, H., "Predicting Parole Success", Journal Criminal Law & Criminology, 1923, 14, pp. 405-413.
3. Burgess, E. W., "Factors determining success or failure on parole", in Bruce, A.A. (Ed.), The Workings of the Indeterminate Sentence Law and the Parole System in Illinois, Springfield, Ill: Illinois State Board of Parole, 1928, pp. 205-249.
4. Glueck, S. and Glueck, E. T., 500 Criminal Careers, New York, Knoff, 1930.
5. Vold, G. B., "Prediction methods and parole," Hanover, N. H.: The Sociological Press, 1931.
6. Glaser, D., "Testing correctional decisions," Journal Criminal Laws, Criminology and Police Science, 1955, 45, pp. 679-684.
7. Manheim, H. and Wilkins, L., Prediction Methods in Relation to Borstal Training, London: Her Majesty's Stationery Office, 1955.
8. Miner, J. B., Personnel Psychology, Macmillan Co., 1969.
9. Guion, R. M., Personnel Testing, New York: McGraw-Hill, 1965.

10. Miner, op. cit.
11. Glennon, J. R., Albright, L. E. and Owens, W. C.,
A Catalog of Life History Items, Washington, D.C.:
Division 14, American Psychological Association, N.D.
12. Cureton, E. E., "Recipe for a cook book," Psychology
Bulletin, 1957, 54, pp. 494-497.

CHAPTER IV

INDIVIDUAL PERSONALITY ASSESSMENT

The Bender Gestalt Test, which is essentially the task of copying nine simple drawings, has had a great appeal to testers and researchers. It is brief, simple, usually attractive to subjects to be tested, and provides a non-threatening introduction to other testing procedures. Most clinicians include it in their battery of diagnostic evaluation, and have considerable confidence in its value as a test of organicity. It is also one of the most widely used tests for psychologists working in correctional settings. There are a number of variations in administration, scoring and interpretation. Predictive validities are low, as in the case of all projective tests, and test-retest reliabilities are also low. Performance on this test has been shown to be affected by a wide range of variables, from the size of the paper to the knowledge of the number of drawings to be copied, for example. In short, it is felt to be a useful clinical instrument, but reliability and validity coefficients are very low.⁽¹⁾ Table 2 shows the few offender studies using the Bender Gestalt.

The Draw-a-Person Test is also one of the most widely used projective instruments in psychological clinics. There are lists of items published to enable its use as an intelligence test or as a projective measure of personality. Different theories of personality have resulted in different methods of scoring or interpreting different aspects of the drawings. Unfortunately,

these are often contradictory, even in the same system. In general, reliability and validity are remarkedly low, when used as a personality measure, although the test may be useful in conjunction with other instruments or other data.⁽²⁾

It should be noted that when the DAP is used with the Harris revision as an intelligence test, split-half reliabilities are in the .70's and .80's.⁽³⁾ Retest reliabilities are slightly lower, ranging in the .60's and .70's. Validity, as determined by correlations with the WISC, WAIS, Stanford-Binet and PMA, are substantial and significant. It should also be noted that as an intelligence test, its use is appropriate only with children.

The House-Tree-Person (HTP) Technique is, as the name indicates a technique or process, rather than a test.⁽⁴⁾ Simply described, it is a projective device which consists of two phases in four steps, intended to accomplish a clinical approach to an analysis of the total personality. In the first phase the subject is asked to make a free hand pencil drawing of a house, a tree, and a person. In the second phase he is again asked to draw a house, a tree, and a person, but this time using crayons. The second step of each phase is his description, definition and interpretation of the drawings he has made, and his associations concerning them.

For the achromatic Phase I there is a set of 60 primary questions for the Post-Drawing Interrogation (P-D-I). For the chromatic Phase II there are 22 formal questions in the P-D-I. However, the P-D-I is not intended to be rigid, and the examiner

is expected to pursue further any line of interrogation or interpretation which appears fruitful.

The HTP may also be quantitatively scored to obtain a measure of intellectual capacity or I.Q. However, HTP is not intended to serve as an intelligence test in the traditional sense of that term. Buck points out that HTP I.Q.s are valuable signposts, but only signposts.⁽⁵⁾ The formal intelligence test is intended to measure only intelligence, in a highly structured situation that does not arouse emotion, and to yield I.Q.s which are highly stable. The HTP I.Q., on the other hand, is almost completely unstructured, is administered in a situation that is intended to arouse emotion, yields I.Q.s which are highly variable, and considers intellectual function as one aspect of an interrelated total personality constellation.

As a projective instrument for the assessment of personality characteristics, HTP has both the assets and the pitfalls of other projective techniques, not the least of which is that it permits the examiner to project his personality into the interpretation of the drawings. Svenson concludes that there is some evidence that the overall drawing may be indicative of organicity or severe pathology, but that individual evaluation is extremely risky and uncertain.^(6,7) On the other hand, as a clinical tool in the hands of a skilled, trained and experienced examiner, the HTP may well yield significant clues to the total personality and present a basis for individual treatment.

The HTP is primarily an individually administered procedure, but it may be used in group administration as well, though with less sensitivity and fewer clues to personality adjustment.

The Thematic Apperception Test (TAT) and the Children's Apperception Test (CAT) both present the individual with rather vague pictures which may be described in an infinite variety of "stories". It is felt that the content of these stories tap fantasy in a way that permits conceptualizations of individual personality, group dynamics, and to theories of personality. There exists a voluminous body of literature concerned with interpretation and significance of a great many aspects of content of response, nature of stimulus, and so on. There are studies of numerous groups, examiner biases, and so on. The emphasis on validation studies seems to have decreased, partly because it was non-productive, and partly because examiners have come to realize that this is not a test, but a method of studying personality.⁽⁸⁾ Offender studies using the TAT are shown in Table 3.

The Rorschach differs qualitatively and quantitatively from the TAT, but is another projective instrument which, though more popular and longer used, is subject to the same criticism and limitations. The stimulus here is a series of ten cards, some black and white as are the TAT cards, and some in color, on which appear inkblots to which the individual responds in unstructured manner. A number of variations have been developed, including a group test, and attempts to structure the response. Several systems of scoring have also been developed, resulting in

numerous inconsistencies and disagreements. Basically, as with the TAT, it is not a test, but a clinical method of studying personality.⁽⁹⁾ Rorschach findings with offenders are found on Table 4.

There are any number of other attempts at the projective study of personality, such as incomplete sentences, drawings of various objects, matching of objects of differing form and color, and so on.

In general, projective tests of personality have low reliability and validity. They are often difficult to administer, are always difficult to interpret, and depend almost entirely on the sophistication, training, and experience of the examiner, plus the availability of other data, such as personal or occupational history, interview, and so on. They have little predictive value.

The basis for the I-Level Classification System was suggested in 1957 in a paper by Sullivan, Grant and Grant and further developed by M. Q. Grant (later M. Q. Warren).^(10,11,12) It has been widely used as a method of classifying offenders, as an aid to differential treatment, and in management and assignment decisions by the California Youth Authority. It has also been used in some institutional and parole programs within the Youth Authority.⁽¹³⁾

The level of integration or I-level is described as "a perceptual frame of reference by which the individual integrates his experience". The individual's philosophy of life is developed from a consistent set of expectations and attitudes. As the

individual develops, so does the frame of reference which integrates his experience.

The I-level classification system originally described seven levels, but the work with delinquent offenders has been essentially limited to levels I₂, I₃ and I₄, as these appear to include nearly all offenders. Within these three levels are found nine subtypes:

<u>Level</u>	<u>Subtype</u>	<u>Name</u>
2	Aa	Unsocialized, Aggressive
	Ap	Unsocialized, Passive
3	Cfm	Conformist, Immature
	Cfc	Conformist, Cultural
	Mp	Manipulator
4	Na	Neurotic, Acting-out
	Nx	Neurotic, Anxious
	Se	Situational Emotional Reaction
	Ci	Cultural Identifier

The I-level classification system as originally developed suffered from the need to train individuals in the use of lengthy, clinical-type interviews as a major diagnostic tool. Training in the procedure requires a five-week course, and weeks of practice following the training. A further problem reported has been the tendency for interviewers to "pull" the offender to the highest level possible.

It seems regrettable that in seventeen years of use and application, classification and study, there is still no report of its use in a predictive validity study. That is, there is no literature following a population of youths over time to see the differential arrest rate by level and type. Numerous offenders have been classified; treated in various ways in many institutions; released, pardoned, paroled, or escaped.

Reporting on this body of data has been scanty. There does not seem to be, for example, any data showing differential success rates on probation by I-level. The studies that do report differential recidivism rates by I-level are compounded by having selected populations, differential treatments and outcome variables not really independent of treatment staff.^(14,15) It is hoped that eventually a relatively large group of subjects will be classified by I-level at an age no greater than 10 years; that this group will be followed for about 10 years, and that differential offense rates by group will be reported. This is necessary before the real ability of the system to predict can be determined.

TABLE 2
STUDIES USING THE BENDER-GESTALT TEST

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
474-1006	Levy, R. & Henning	'70	224		Report of computer use in diagnosis. The B-G information is programmed in.
474-3390	Matranger, J. et al	'72		Male, Negro Offenders	Normed Bender-Gestalt for I.Q. levels and Race.
474-3871	Svenson, W. & Grimes	'58	451	Sex Offenders	A simple study of personality types of sex offenders. Not hypothesis testing.
474					

TABLE 3
STUDIES USING THE THEMATIC APPERCEPTION TEST
(TAT)

REFERENCE NOS.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
674-4636	Shore, M. et al	'64	20	Anti-social boys	Used TAT to show changes in guilt before and after psychotherapy.
474-3858	Silver, A.	'63	80	Delinquents vs. Non-delinquents	Compare MMPI to TAT results. MMPI predicts they question use of TAT.
474-3162	Young, F.	'56	68	Delinquents	Simple description of population as given by TAT responses.

STUDIES USING THE RORSCHACH TEST

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
574-4373	Argyle, M.			Juvenile Delinquents	Not a study, but a review of studies for classification.
474-3766	Kahn, M.	'59	39	Murderers & Burglars	Compared 2 groups on Rorschach protocols.
574-5153	Perdue, W. & Lester, D.	'72	30	Rapists	Rapists do not differ significantly from other violent (non-sexual) offenders.
474-3849	Rudie & McGaughan	'61	311	Alcoholics	Study differentiates between essential and reactive alcoholics.
474-3871	Svenson, W. & Grimes	'58	45	Sex Offenders	Description of Personality of sex offenders.

CHAPTER IV
REFERENCES

1. Kitay, P. M., "A review of the Bender-Gestalt test," in Buros, O. K. (Ed.) The Seventh Mental Measurement Yearbook, New Jersey: Gryphon Press, 1972, p. 161.
2. Harris, D. G., "A review of the Draw-a-Person test," in Buros, O. K. (Ed.) The Seventh Mental Measurement Yearbook, New Jersey: Gryphon Press, 1972, p. 165.
3. Anastasi, Anne, "A review of the Goodenough-Harris Drawing test", in Buros, O. K. (Ed.) The Seventh Mental Measurement Yearbook, New Jersey: Gryphon Press, 1972, p. 352.
4. Buck, John N., The House-Tree-Person Technique: Revised Manual, Los Angeles, California: Western Psychological Services, 1966.
5. Ibid
6. Svenson, Clifford H., "Empirical evaluations of human figure drawings," Psychological Bulletin, 1957, 54, pp. 431-466.
7. Svenson, Clifford H., "Empirical evaluations of human figure drawings," Psychological Bulletin, 1968, 70, pp. 20-44.
8. Dana, R. H., "A review of the Thematic Apperception test", in Buros, O. K. (Ed.) The Seventh Mental Measurement Yearbook, New Jersey: Gryphon Press, 1972, p. 181.

9. Burstein, A. G., "A review of the Rorschach," in Buros, O. K. (Ed.) The Seventh Mental Measurement Yearbook, New Jersey: Gryphon Press, 1972, p. 175.
10. Sullivan, C. E., Grant M. Q., and Grant, J. B., "The development of interpersonal maturity: application to delinquency," Psychiatry, 1957, 20, pp. 373-385.
11. Warren, M. Q., Interpersonal Maturity Level Classification: Juvenile, 1961, California Youth Authority, Division of Research.
12. Warren, M. Q., "The case for differential treatment of delinquents," Annals American Academy Political Social Science, 1969, 381, pp. 47-59.
13. Warren, M. Q., Interpersonal Maturity Level Classification: Juvenile, "Diagnosis and treatment of low, middle and high maturity delinquents," 1966, California Youth Authority.
14. Palmer, Ted, "The Youth Authority's Community Treatment Project," Federal Probation, March 1974, pp. 3-14.
15. cnf. supra and Palmer, Ted, "Matching worker and client in corrections", Social Work, March, 1973, pp. 95-103.

CHAPTER V
GROUP PERSONALITY ASSESSMENT

The Minnesota Multiphasic Personality Inventory (MMPI) is generally recognized as the foremost instrument in the field of objective clinical assessment. It consists of some 550 different items to be answered as "true" or "false", items such as, "I don't seem to care what happens to me," "I like dramatics," and "most people are honest chiefly through fear of being caught." It is available in three forms - a card sorting form, a booklet form, and a computerized version. In this last, a system has been devised that will print out descriptive paragraphs to match scale scores.

The items are scored in clusters or scales, each of which offers a score on one of nine separate dimensions of behavior or traits, each named after the pathological conditions it was designed to detect - Hypochondriasis (Hs), Depression (D), Hysteria (Hy), Psychasthenia (Pt), Schizophrenia (Sc), Hypomania (Ma), Psychopathic deviate (Pd), Masculinity-femininity (Mf), and Paranoia (Pa). In addition there are three "validity" scales to check the genuineness of the responses - the F Scale or "validity" Scale which measures the tendency to answer items in an atypical fashion, the L or "lie" Scale designed to identify deliberate efforts to distort responses, and the K Scale intended to correct for response bias introduced by social desirability. The ratio of F to K is also considered a measure of faking.

While the scale scores may be used separately, the usual custom is to construct a "profile" by displaying the scores on each scale side by side on a graph oriented around a base line equal to the mean of each trait. The resulting profile enables the clinician to see which scale scores are elevated or depressed and to what extent, and to note any significant interrelationships, such as the tendency for the mutual appearance of elevated scores on F, Pd, Pa, Sc, and Ma when testing management problems among offenders.

The MMPI was intended as an open-ended instrument. Of the 550 items, only 399 are used for the standard clinical scales developed originally, plus the Social I-E scale of Drake. The other 151 items are used for the development of new scales, and for research. The Form R booklet is so arranged that, without disturbing the content of the item, only the first 399 need to be used to obtain all of the standard scale scores, thus simplifying administration.

The original MMPI was individually administered, using separate cards for each item. The response to each item was recorded on a separate answer sheet or record form. The authors have preserved these data of the original standardization group, as well as for many other groups. Consequently, the numerous new scales developed can be compared with the original standardization sample, or cross-validated with other groups of known characteristics, thus providing for a continuity of research and development that is nearly unique in testing history.

None of the original scales used all 399 items, and show considerable variety in the range of number of items used. In general, the newer scales developed tend to be shorter than the original scales, and thus to suffer from somewhat limited reliability, but this is not necessarily so. As in the case of any test, the user must take reliability and validity of each scale into consideration in the selection of a specific instrument, or in the interpretation of results.

The original purpose of the MMPI was to evaluate major personality characteristics that are commonly associated with disabling psychological abnormality. The original scales have since been found to have meaning within the normal range of behavior. The scales were constructed contrasting the responses of a sample of approximately seven hundred individuals who visited the University of Minnesota Hospitals, were considered normal, and to be a representative cross section of the Minnesota population; with the responses of over eight hundred carefully studied patients from the neuro-psychiatric division of the same hospitals.

The chief criterion was prediction or agreement with diagnosis of the neuropsychiatric staff, rather than statistical measures. Since then, however, literally hundreds of studies have been published, reporting numerical estimates of reliability and validity of profile analysis or single scales for innumerable kinds of groups under all kinds of situations, and using a wide variety of criteria. The serviceability and usefulness of the instrument appears to be soundly established.

More than 60% agreement is quoted in the Manual and where the elevated scale score did not agree with the diagnosis, the trait concerned still was noticeable in the clinical picture. It is interesting to note that 30 years later in trying out a computerized system with automatic descriptive printout, Lachar found that with some 107 paragraphs appearing 7,555 times and checked by each patient's individual therapist, the paragraphs were judged invalid only 9.7% of the time.⁽¹⁾

The MMPI has been translated into numerous languages, including Spanish. Thus, research studies may include the Puerto Rican and Chicano offenders.

Thus, the findings of Monachesi and Hathaway in 1969 pointing out the elevation of Pd, Sc, and Ma in delinquency,⁽²⁾ is echoed in 1974 by Flanagan and Lewis in comparing absolute first offenders with those with previous records of juvenile delinquency.⁽³⁾ In all honesty, however, they point out that the large size of their sample inflates the statistical significance of what is after all, a small absolute difference, and question its practical value while admitting its research and theoretical interest. Gregory also reports a study replicating the actuarial correlates of three MMPI code types and with a sequential stepwise regression strategy comes up with the classical three types of delinquent - Psychopathic, Adjusted, and Neurotic.⁽⁴⁾ Perhaps the time has come to abandon the theoretical approach to the MMPI and recognize its practical utility.

As we have said above, the multiplicity of items (550) on

the MMPI offers the opportunity of developing specialized scales for special populations. The resulting proliferation of special scales has hit the corrections field with its peculiar problems of recidivism, prison adjustment, escapism, prediction of type of crime, and control of violence. All of these have led to the development of special scales on the MMPI.

The large number of items included in the MMPI tends to lead the casual observer to believe that all of the items are used. We have already noted that only 399 items are used for the original scales. The number of items actually used in obtaining scores on the various subscales ranges from 15 on the deception scale to 78 on the schizophrenic scale. It should not surprise us that various authors report reliability coefficients for the various scales ranging from the .40's to the .90's.

Cottle for example, reported reliability coefficients ranging from .46 for the 15 item Lie scale to .91 for the 60 item feminine version of the Masculinity/Femininity Scale.⁽⁵⁾ Any subscale must present its own evidence of its reliability and validity, and cannot depend upon the reliability or validity of the primary instrument. As we have already noted, the shorter the subscale the more likely it is to suffer from limited reliability, and the more cautious must be the user in applying the scale.

Also, when attempting to differentiate a special population from the larger population from which it was drawn, i.e., escapees from a general, non-escapee prison population, we usually are

forced to deal with smaller groups, i.e., escapees are much rarer than prisoners in general, and the resulting differentiation may be a peculiarity of the small experimental sample, i.e., we do not have enough escapees to assure a representative, random sampling of the type. Again this increases the need for further cross-validation and replication, and lucky indeed is the researcher who does not find a lowered reliability and validity with his second testing. For this reason many subscales that appear promising on first derivation, disappear with further investigation.

In 1968, Kincannon offered an abbreviated form of the MMPI consisting of only 71 items, the Mini-Mult.⁽⁶⁾ Subsequent investigation (i.e., Armentrout and Rouzer, 1970⁽⁷⁾; Mlott, 1973⁽⁸⁾, Hobbs, 1974⁽⁹⁾) indicates the usefulness of the shorter form when used with caution in view of its limited reliability and informational potential.

Many of the newer scales have some value when used with caution. We will deal with some of these later in discussing special correctional problems such as escapism, recidivism, and tendency to violence.

Panton tried in 1958 to obtain differential profile configurations according to type of crime committed (white collar, aggravated assault, robbery, property theft, aggravated sex perverse) with no success.⁽¹⁰⁾ In 1974 Christensen and LeUnes replicated Panton's work using a more complicated statistic, were able to separate out murderers and addicts.⁽¹¹⁾ The Prison Adjustment Scale, however, did not work for recidivism. Subsequent

studies confirm the inability of MMPI scales to successfully detect recidivism, i.e., Wattson⁽¹²⁾, Mack⁽¹³⁾ and Jaman⁽¹⁴⁾. Nor does it work with traffic offenders, Whinery⁽¹⁵⁾. One suspects that recidivism is a function of much more than simple personality structure and that any MMPI scale used for this purpose would need supplementary information of a different sort.

Shupe and Bramwell⁽¹⁶⁾ replicate the earlier work of Beall and Panton⁽¹⁷⁾ on an "escape" scale (Es). Again, some differentiation seems possible, but not at the level of practical application, although Shupe and Bramwell in discussing hits, misses, and false positives do say the ultimate decision to use the scale must depend upon what the system will stand for.

In the prediction of violence, where a scale would fill a crying need, the results have been negative, at any but a research level (Jaman, et al⁽¹⁸⁾, Wenk et al⁽¹⁹⁾) and we can agree with Megargee and Mendelsohn⁽²⁰⁾ that "Certainly the data indicate it is difficult if not impossible to identify an assaultive individual with reasonable accuracy using the MMPI scales now available." As both Blackburn⁽²¹⁾ and Deiker⁽²²⁾ point out, however, the problem needs theoretical clarification, and future research may improve the situation.

In summary, although certain aspects of the development, standardization and validation of the MMPI have been questioned by various authors, the experience of many authors has shown it to be a highly useful instrument for the study of personality characteristics in normal populations as well as in disturbed groups.

It presently predicts some types of behavior more successfully than others. In the studies of various classes of offenders it has not been as successful as many have hoped it would be. These failures may be because MMPI was not originally designed to measure many of the traits to which it has been applied or to the extremely variable nature of human behavior, to the research designs employed, or to some other cause or causes. Nevertheless, it appears to be a promising instrument, and the failures of the past should not keep us from the future, hopefully better and more successful, studies. The numerous studies using the MMPI are shown on Table 5.

The 16PF Questionnaire is an objectively scored test intended for use with individuals 16 or over.⁽²³⁾ Forms A & B have 187 items, forms C & D have 105 items. The test is easily administered and easily scored. The test is the product of more than 30 years of factor analytic research, and measures 15 personality variables in addition to intellectual ability. Scores are reported in profile form in stens (ten equal interval standard score points) which can easily be translated into centiles or stanines.

The traits measured are:

- | | |
|---|---|
| A - Reserved vs. outgoing | L - Trusting vs. suspicious |
| B - Less intelligent vs.
More intelligent | M - Practical vs. imaginative |
| C - Affected by feeling vs.
Emotionally stable | N - Forthright vs. shrewd |
| E - Humble vs. assertive | O - Placid vs. apprehensive |
| F - Sober vs. happy-go-lucky | Q1- Conservative vs. experimenting |
| G - Expedient vs. conscientious | Q2- Group dependent vs.
self-sufficient |
| H - Shy vs. venturesome | Q3- Undisciplined self-conflict
vs. controlled |
| I - Toughminded vs.
tenderminded | Q4- Relaxed vs. tense |

Construct validity from ten successive factor analyses ranges from .41 to .94 for different scales and forms. Obviously, some of the scales are too limited to be useful in predicting individual behavior. Reliability ranges from .45 to .90 for test-retest measures and from .71 to .93 for split-half correlations.

Although the author claims the 16 factors are independent, some other studies question the independence of the scales; Becker suggesting there are really only 8 factors,⁽²⁴⁾ and Adcock reporting that though the factors may be independent, they are not necessarily uncorrelated.⁽²⁵⁾

Two studies have shown the 16 PF to be superior to other instruments for classification and in distinguishing between subgroups, such as between serious and non-serious offenders.^(26, 27) None of the instruments studied, however, successfully predicted institutional adjustment. In addition to being much shorter than

the other instruments used, the 16PF showed significant agreement with ratings by clinicians, suggesting that it has some advantage as a large scale initial screening device.

The extensive literature related to this inventory support Adcock's⁽²⁸⁾ prediction that the 16PF may eventually become the standard for selection. At present, however, this inventory, like the others, suffers from a lack of well-planned and well-executed studies of predictive validity, specifically in the studies of delinquents and other offenders. Offender studies using the 16PF are shown on Table 6.

The Jr.-Sr. High School Personality Questionnaire (HSPQ) is intended for the age range 12 through 17, just below the range of the 16PF. It covers fourteen factors or source traits, identified and referred to by letters of the alphabet, keeping the same designations as those used in the 16PF. Two of these do not appear in the 16PF, and four from the 16PF have not been included in the HSPQ. The questionnaire is easily scored in a few minutes.

Each scale contains only ten items, so that reliability has been low, over a period of time, when only one form was used. However, if two forms are used, reliability over a period of one year is .63.

This inventory has been seriously challenged by many in the field as deficient in evidence of validity.^(29,30) The test does not seem to be widely used, in any event, and users should approach interpretation with caution. As a research instrument it appears to offer more value than as a practical tool for

classification.^(31,32,33) Research studies using the HSPQ are shown on Table 7.

The California Psychological Inventory (CPI) consists of 480 items, twelve of them duplicated and 20 of them not scored at all. Approximately 200 of the items were taken directly, or slightly adapted, from the MMPI, which raises some question of the significance of the correlations between the two tests. The CPI is a self-administered inventory using a separate answer sheet and requiring about 45 minutes of test time.

Like the MMPI, the CPI was intended as an open-ended instrument. That is, it was expected that new scales would be developed from time to time by the author, and by others. Although some others have developed scales (Goodstein and Schroeder⁽³⁴⁾, a managerial key; Jogan⁽³⁵⁾, an empathy scale, and Leventhal⁽³⁶⁾, an anxiety scale), these are not relevant to a study of offenders. The author has himself apparently preferred to use regression equations to predict various outcomes. Others, also, have tended to follow this same pattern (e.g. Kirk, Cummings, and Hackett⁽³⁷⁾ in studies of dental students.)

The inventory provides 18 scales intended to develop descriptive concepts of broad personal and social relevance. The inventory is primarily intended for use with normal subjects, measuring characteristics important for social living and social interaction. Although it has been found to be useful in evaluating some problem groups, it has been most widely used and is expected to be most useful in schools, business and counseling settings

dealing primarily with socially functioning individuals.

The 18 scales are grouped into four categories, bringing together for interpretational purposes scales with similar implications.

Class I - Measures of poise, ascendancy, self assurance and interpersonal adequacy:

- | | |
|-----------------------------|-----------------------------|
| 1. Do - Dominance | 4. Sp - Social Presence |
| 2. Cs - Capacity for Status | 5. Sa - Self Acceptance |
| 3. Sy - Sociability | 6. Wb - Sense of Well Being |

Class II - Measures of socialization, maturity, responsibility, and interpersonal structuring of values:

- | | |
|------------------------|--------------------------|
| 7. Re - Responsibility | 10. To - Tolerance |
| 8. So - Socialization | 11. Gi - Good impression |
| 9. Se - Self Control | 12. Cm - Communality |

Class III - Measures of achievement potential and intellectual efficiency:

- 13. Ac - Achievement via Conformance
- 14. Ai - Achievement via Independence
- 15. Ie - Intellectual efficiency

Class IV

- 16. Py - Psychological - mindedness
- 17. Fx - Flexibility
- 18. Fi - Femininity

Although each scale may be used independently, it is intended that the individual be evaluated by interpretation of the profile of the 18 scales, and their interaction. As the scales

are intended to identify individuals who will behave in a certain way and be described in a certain manner, this interaction of high and low scores on the several scales is important. A weakness of such interpretation, as on the MMPI, EPPS, and other similar instruments, is that the validity of the interpretation is directly affected by the general psychological sophistication of the interpreter and his experience with the instrument.

Test - retest reliabilities are reported for a group of 200 male prisoners retested at one to three week intervals, and for 101 male high school students and 125 female high school students, retested after an interval of approximately one year. Correlations ranged from .48 to .74 for the students, with median retest reliabilities of .65 for the males and .68 for the females. As might be expected because of the shorter retest interval, the correlations for the prisoners are higher, ranging from .49 to .87 with a median of .80.

Validity has been determined primarily by the method of development of the scale, and by correlating specific scale scores with external social behavior criteria. Thus, Ac, Ai and Ie have been correlated with school grades and/or measures of intellectual power; psychological mindedness (py) has been estimated by observing the relationship of mean scores on this scale obtained by various groups presumably varying in this trait. The Socialization scale has been validated in the same way, including such disparate groups as "bank officers" and "prison inmates".

A good many others have used the inventory, and especially

the Socialization scale, in a wide variety of studies ranging from distinguishing delinquents and prisoners from non-offenders, and developing typologies of offenders on the one hand, to the probability of becoming delinquent, or predicting recidivism on the other. There is also some evidence from the Community Treatment Project in California, and a study of traffic offenders in Oklahoma, that CPI can be helpful in selecting offenders who will respond positively to different types of treatment.

In short, the Socialization scale (So) of the CPI has been found to consistently distinguish between delinquent and non-delinquent groups. (38, 39, 40, 41, 42, 43, 44) Little attempt has been made to predict susceptibility to differential treatment, or to predict recidivism, but the few attempts that have been made indicate that the So scale can probably be a highly effective and parsimonious instrument for such purposes. (45, 46, 47, 48) Table 8 shows studies using the CPI.

Jesness has proposed the use of the Jesness Inventory to objectify and simplify the method of determining the I-level. (49) (See p. 34, Chapter IV for a discussion of the I-level). The computer scoring system of Consulting Psychologists Press must be used for this approach. However, as this results in ambiguity in about 46% of the cases, the final classification for the I-level requires the use of the Jesness Behavior Check List. The use of the BCL reduces ambiguity to about 27%.

The Jesness Inventory which is used to obtain the I-level classification is a 155 item scale that provides scores on eleven

personality characteristics.⁽⁵⁰⁾ Three of the scales are empirically derived and seven are derived from cluster analysis. The eleventh scale is the Asocial Index, based on the entire inventory, and is used to predict delinquency.

The empirical scales are:

SM - Social Maladjustment IMM - Immaturity

VO - Value Orientation

The cluster scales are:

Au - Autism SA - Social Anxiety

Al - Alienation Rep - Repression

MA - Manifest Aggression Den - Denial

Wd - Withdrawal

The Asocial Index was derived by the discriminant function, with the general equation derived from scores of 963 delinquent and 925 non-delinquent boys. The formula works equally well with girls, achieving 84% accuracy. However, a separate discriminant analysis for the female sample yielded 86% accuracy.

Jesness reports reliability coefficients based on odd-even items for 1862 boys ranging in age from 10 to 18, and test-retest data for 131 boys. Odd-even correlations ranged from .45 to .79 uncorrected, with a median of .56. Corrected by Cronbach's formula, coefficients ranged from .62 to .88 with a median of .71. With the test-retest group the uncorrected coefficients ranged from .35 to .67 with a median of .57. Corrected, there were .40 to .79 with a median of .69. However, no reliability data are provided for the Asocialization scale.

Validity has been determined first by differences in mean scores for selected groups of delinquent and non-delinquent samples. The scales have also been correlated with the subscales of the CPI. The subscales are significantly correlated, ranging from $-.76$ for Achievement via Conformance and Social Maladjustment to $-.42$ for Achievement via Independence, for example. Interestingly enough the highest positive correlation reported is $.71$ between the Asocial scale and Femininity, thus raising questions about the validity of each of these scales.⁽⁵¹⁾

The Jesness Inventory is easily administered, requiring approximately 45 minutes for group or individual administration. It may be administered using a separate answer sheet with either tape recorder or booklet. Scoring requires eleven stencils, and a profile sheet. The Asocial Index is obtained by computation, using weighted scores and a formula. The final T scores are obtained from a table. As already noted, a computer scoring service is necessary when scoring the inventory to obtain I-levels.

Cowden, Peterson and Pacht found that the Jesness Inventory's Asocial Index did not effectively discriminate between delinquents having good or poor institutional adjustment. They did find that it did a more effective job than the MCI as a screening and classification instrument in discriminating among subgroups.⁽⁵²⁾ Cowden et al, also found that there were significantly higher scores on some scales for poor adjusters as compared to good adjusters.

Griffiths reports that the Fremont program showed that a decrease in the Alienation score during the program was associated

with subsequent parole success, and an increase in Alienation was associated with parole failure.⁽⁵³⁾ Weintraub concludes that the inventory identifies but does not predict delinquency, and is of limited usefulness.⁽⁵⁴⁾ Fraas and Price report failure in attempting to predict recidivism in AWOL soldiers,⁽⁵⁵⁾ even though the groups had statistically significant differences.

Kelly and Baer also found some short term usefulness in the Jesness Inventory when administered to 60 male delinquents before and after participation in an Outward Bound Program.⁽⁵⁶⁾ The Social Maladjustment, Value Orientation, Autism, Alienation, Manifest Aggression and Repression Scales showed significant changes in the direction of more favorable social attitudes. No evidence is presented to show that these changed attitudes were associated with changes in behavior.

The I-level typology derived from interview data has not been used by Warren et al to predict parole success, but it does appear to distinguish those groups committing more serious violations and violent violations.⁽⁵⁷⁾ Butler and Adams applied a Q sort to the I-level items, but were not able to objectify the typologies.⁽⁵⁸⁾ Although their study leads them to question the therapeutic suggestions that follow from I-level theory, they feel that the classification has merit for differential treatment.

Jesness reports that, when placed in living units according to I-level subtype, there were fewer behavioral problems and rule infractions, and unit management problems were decreased. Still, the experimental program had no noticeable effect on parole performance.⁽⁵⁹⁾ However, Warren reports that experimental groups

who were well-matched with their caseworkers in the Community Treatment Program, had a much lower failure rate than an experimental group who was not well-matched.⁽⁶⁰⁾ She points out that the issues are complex, and involve more than the classification of the offender. The selection of specific offenders to receive specific treatment in specific settings to achieve specific goals are complexly interrelated with each other, and with community attitudes as well.

In summary, the Jesness Inventory and I-level typology appear to be significantly helpful in classification of offenders for management and helpful in institutional handling of offenders. Although the schema appears to be intriguing, there is little to show that either classification or differential treatment has been of significance in affecting parole outcomes.

The Tennessee Self Concept Scale (TSCS) consists of 90 items comprising the self concept scale, plus 10 items taken from the Lie scale of the MMPI which provides the self-criticism score. The latter is a measure of defensiveness, and serves as an indication of the validity of the responses. There are two forms, Counseling, and Clinical and Research. The booklet and test items are the same for both, but the former, which is quicker and easier to score, yields only 14 scores, while the latter yields 30 scores, and is considerably more difficult to score and to interpret. As the Counseling form is recommended by Fitts for use where the results are to be interpreted to the counselee, or where the examiner is lacking in sophistication in psycho-pathology,

this review is concerned only with the Counseling form. However, a well trained counselor may well prefer to use the Clinical and Research form, especially if he is engaged in some research project. Administration time is generally less than 20 minutes.

The scores obtained are:

1. Self criticism (SC) score: a measure of defensiveness or openness.
2. Positive Scores (P):
 - a. Identity: "that is what I am"
 - b. Self satisfaction: "this is how I feel about myself"
 - c. Behavior: "this is what I do"
 - d. Physical self: his view of his physical self
 - e. Moral/ethical self: self description in a moral-ethical frame of reference
 - f. Personal self: his sense of personal worth
 - g. Family self: his feeling of value as a family member
 - h. Social self: reflects his sense of adequacy as he interacts socially

Three other scores are also reported. The Variability (V) score measures the consistency of the responses, and the Distribution (S) score reflects the degree of certainty with which the individual has responded. A Time score is also recorded, a record of the length of time required to complete the inventory, but there is no information regarding its significance.

Richards, Mates and Whitten report a study of 102 girls

aged 12 to 18 confined to state correctional schools in Florida and Tennessee.⁽⁶¹⁾ They found that the two samples apparently came from the same population of delinquent girls. They tended to see themselves in the same way that they see man in general. If they were positive about themselves they tended to be positive about human nature. Those with minimal personality disorders are more favorable in their attitudes than those who are more disordered. In general, this group differed from a normal population in presenting greater self-conflict, lower self-satisfaction, high behavior variability, low personality integration and inadequate social-family-religious life.

Fitts has reported several studies of his own and of others, which are summarized in his monograph, "The Self Concept and Delinquency". He finds that delinquents differ from non-delinquents in self concept, and that those who are more delinquent (that is, habitual delinquents or criminals) have self concepts that differ from first offenders. He concludes that the self concepts of delinquents are "more negative, more uncertain, more variable, and more conflicted. They are also less defensive, show strong acquiescent tendencies, simply much pathology and little personality integration. Delinquents are down on society and often in conflict with society, but it seems safe to conclude that they have the same difficulties with themselves."⁽⁶²⁾

STUDIES USING THE MMPI
TABLE 5

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
474-4010	Armentrout, J.	'70	125	delinquents	Compares mini-mult to MMPI. Does almost as well.
574-4181	Beall, H. & Panton	'56	503	Prisoners	Develops recidivism scale, misclassifies 1/4 cases.
474-3181	Blackburn	'71	56	Murderers	Shows 4 types of murderers, as measured by MMPI Scales.
474-3400	Blackburn, R.	'72	165	Psychiatric maximum security patients	Used MMPI to study personality of psychiatric offenders.
474-3167	Brandt, J. & Hood	'68	915	College males	Relation of MMPI to Strong Vocational Interest Test.
474-3859	Briggs et al	'61	489	Delinquent boys	Prediction of delinquency by MMPI and others.
574-4351	Caldwell	'59	495	White & negro male offenders	Developed criminal types by MMPI & offense history.
474-3847	Capwell	'45	189	Delinquent girls & controls	Developed personality profiles - Showed MMPI Scales can discriminate
574-4404	Deiker, T.	'74	40	Property offenders vs. offenders against persons	Validation of MMPI violence scales.

TABLE 5 (cont.)

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
574-4257	Edwards, J.	'63	44	Prisoners	Used MMPI to define personality of prisoners with poor institutional adjustment.
474-3200	Eysenck	'69			Review of literature & compares MMPI to MPI.
474-3170	Fisher, G.	'67	784	Prisoners Black & white	Used MMPI to study Marlowe-Crowne Social Deviance Scale
474-3853	Fry	'49	443	Male/Female Prisoners and college students	No hypotheses
674-5243	Gendreau P. & Gend.	'70	133	Heroin addicts	Delineation of addict prone personality.
474-3395	Godfrey	'72	178	Male felony offenders	MMPI part of Battery I "post-dict" 3 types of offenders
574-4230	Gough, H., et al	'65	739	CYA parolees	Predicted recidivism by CPI, MMPI & B.E.; B.E. best alone; CPI better than MMPI.
474-3848	Gynther, M.	'62	190	White male criminal	F Scale, compared various offenders
474-3371	Gynther et al	'73			An automated MMPI Interpretation.
474-4117	Hathaway & Monacheri et al		1329	Delinquent girls	result reevaluation of MMPI to predict delinquency

TABLE 5 (cont.)

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
574-4346	Hathaway et al	'59	11,000	High School boys & girls	Shows MMPI can be used to predict future delinquency.
474-3872	Hathaway & Monachesi	'51	4000	9th grades	Part of the big MMPI study on delinquency.
574-4333	Hathaway & Monachesi	'57	1958	boys	Post-dictive study to see what scales predicted delinquency.
474-3869	Hill et al	'62	399	Narcotic addicts & alcoholics	Similarities exceed differences. types; psychopaths, primary psychopaths, & neurotic depressed.
474-3846	Hill	'60	270	Narcotics Addicts	
574-4320	Hindelang	'63	582	Adolescent boys	Shows relation of certain personality clusters to reported crimes.
474-0024	Jaman, D.	'69	500	Male prisoners	Shows association of MMPI with Prison Behavior scores.
674-5002	Jaman, D.	'66	244	Violent offenders	An attempt to develop MMPI predictions for violence.
474-0028	Jessness, C. F.	'73	3197	delinquents	Validation of Jesness Inventory vs. MMPI & CPI.
474-3372	Johnson, V. & Cooke	'73	235	Prisoners	Alcoholic Differential Scale, Escape scale, Recidivism scale (evaluate)
474-3854	Jurjevich	'63	170	delinquent girls	Provided normative data on MMPI

CONTINUED

2 OF 4

TABLE 5 (cont.)

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
474-4050	Kanun	'60	504	delinquents	Use of L & F Scales to predict delinquency.
474-3862	Kingsley	'60	100	Psychopaths or others	MMPI Profiles on 3 groups.
474-4008	Kish, G., et al.	'69	216	Alcoholics vs. others	Tests Quay's hypothesis of sensation seeking in alcoholics, rejects hypothesis.
474-3399	Krauss, H. et al	'72	30	Psychopath & non-psychopath prisoners	Study of risk taking in 2 groups.
674-4632	Lawton	'65	30	Prisoners	Subjects could <u>not</u> manipulate scale to test "good".
474-3183	Mack, J.	'69	159	Delinquent boys	MMPI is not good prediction of recidivism in delinquent population
474-4024	Mandel N. & Barron	'66	372	Recidivists	MMPI has <u>no</u> predictive value for recidivism.
574-4336	Monachesi	'50	485	Delinquents & controls Male & Female	MMPI differentiates better between female delinquents & non-delinquent than males.
574-4337	Monachesi	'50	441	Male delinquents & controls	Can differentiate between Institutionalized & non-institutionalized delinquents, chiefly on PI Scale.
574-4338	Monachesi	'48	295	Delinquents & non-delinquents	Personality characteristics that distinguish delinquents from non-delinquents.

TABLE 5 (cont.)

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
474-3873	Morrice, J.	'57	24	Recidivists	
474-3861	Panton, J.	'62	100	Habitual Offender	
474-3863	Panton, J.	'59	2374	Male Prisoners	Simple description of prisoner profiles.
474-3865	Panton, J.	'58	1313	Prisoners	Describes MMPI profiles for differ offenders.
474-4118	Panton, J.	'68	2198	Prisoners	Development of a Parole violation Scale.
574-4201	Panton, J.	'58	381	Prisoners	Differentially identifies 79-88% of adjusted & unadjusted prisoners
474-3416	Peterson, D. et al	'61	406	Delinquents controls (male)	Developed 3 types of delinquents: Neurotic, Delinquent background and psychopath.
474-3383	Platt	'72	183	Male delinquent youths	Compares MMPI to Minimult. Little correspondence between 2 forms and Minimult was little help.
474-3857	Rimple	'58	1802	Boys	Used school data and MMPI to predict delinquency.
474-3860	Rowley, V. & Stone	'62	120	Delinquent & emotionally disturbed boys	MMPI profiles do not distinguish between emotionally disturbed & delinquent boys.

TABLE 5 (cont.)

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
474-3870	Shipe, D.	'62	195	Retarded mental patients	An attempt to measure escape proneness.
474-3201	Shupe, J. & Bramwell	'63	76	Prison escapees	MMPI Escape Scale gives too many false positives.
674-5796	Shupe, J. & Bramwell	'63	76	Prison escapees	MMPI Escape Scale gives too many false positives.
474-3858	Silver, A.	'63	40	Delinquents & orphans	TAT does not do as well as MMPI in discriminating. Study does not give statistics.
74 474-4047	Stanton, J.	'56	200	Prisoners & control	White & negro prisoners have same general MMPI profiles; controls are difficult.
574-4135	Stein	'68	971	Male prisoners	
574-4234	Stein et al	'71	346	Delinquent boys	Uses MMPI & others to delineate typology of delinquency.
574-4223	Stoffer	'69	10	Female narcotic addicts	Reports on 10 female addicts.
474-3864	Stone, F. & Rowley	'62	120	Delinquent & emotionally disturbed girls	Profiles are similar.
474-3369	Sutker, P.	'73	204	Prisoners, addicts, controls	Compared personality profile of 3 groups.

TABLE 5 (cont.)

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
474-3871	Svenson et al	'58	45	Sex offenders	Simply delineates personality characteristics.
574-4156	Svenson and Stumes		45	Sex offenders	Seems descriptive of schools and social history
574-4156	Taylor	'67	33	Delinquent girls	MMPI did not do as well as Group Predictors.
674-5177	Thorne, G.	'71	305	Male & female delinquents and prisoners	Use of the sensation seeking scale.
75 474-3906	Tsubouchi, I. & Jenkins	'69	100	Delinquent boys	Delineates 3 types of delinquencies runaway, socialized, unsocialized aggressive.
474-3194	Wattron	'63	589	Delinquent girls	
474-3192	Wattron	'63	450	Maladjusted prisoners, parolees, recidivists	Constructed a prison maladjustment MMPI Scale that selected 83% of maladjusted.
574-4334	Webster, A.	'54	650	Prisoners	Personality types developed for prisoners.
574-4196	Wirt, R. & Briggs	'59	1958	Young men	Study of those whose outcomes were autonomous in light of their earlier MMPI scores.
574-4176	Whinery, L.	'71	214	16-18 year old traffic offenders	One of the few studies of effects of sentencing traffic offenders. MMPI does not predict recidivism.

TABLE 6
STUDIES USING THE 16 PF

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
574-4202	Cowden, J. et al	'68	134	Delinquent Boys	16PF better than MCI for screening battery, neither predicted behavior adequately.
574-4210	Cowden, J. et al	'71	143	Delinquent boys	16PF more useful than CPI for mass screening; CPI for auxiliary screening in special cases.
574-4213	Cowden, J. et al	'70	156	State Reformatory Inmates	16PF better screening device than MCI, but MCI better predictor of institutional adjustment.
574-0019	Levy, J.				Not a study. Report of Illinois Dept. of Corrections Computerized Test Battery.
574-4273	Palmer, T.	'73			Description of the Georgia Dept. of Corrections computerized psychological assessment.
474-3189	Pierson, G. et al	'67	338	Male Delinquents	Study of personality characteristics.
574-4142	Taylor	'67	33	Delinquent Girls	MMPI did not do as well as Group Predictions.
574-4178	Whinery, L.	'73	214	Traffic Offenders	A long term report on traffic offenders.

TABLE 7
STUDIES USING THE HIGH SCHOOL PERSONALITY QUESTIONNAIRE
(HSPQ)

REFERENCE NOS.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
474-3997	McQuaid, J.	'70	780	Scots delinquent & non-delinquent	showed HSPQ distinguished personality types & delinquents & non-delinquents in Scotland.
474-3189	Pierson, Mosely & Olson	'67	338	Male delinquents	Shows HSPQ's ability to predict delinquency.
674-5272 77	Richards, W., Mates & Whitten	'69	102	Delinquent girls	Little difference in personality structure between delinquent & non-delinquent girls. Institutionalization may not help them.
574-4245	Tyler, V.	'71	168	Male delinquents	
474-3999	White, W., Porter, T.	'70	60	White male delinquents	An attempt to see how a delinquent views the world.

TABLE 8
STUDIES USING THE CALIFORNIA PSYCHOLOGICAL INVENTORY

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
574-4210	Cowden, J.	'71	143	Delinquent Boys	Found 16PF more useful for mass screening; CPI for auxiliary in special cases.
574-4293	C.Y.A.	'66	620	CYA Delinquents	Used CPI to describe population.
574-1952	Dinetz et al	'62	73	Delinquent Boys	Follow up of earlier tests - Good stayed good - bad stayed bad.
474-3855	Gough	'60	20856	Good Citizens, Delinquent Prisoners	Validation of So Scale.
78 574-4230	Gough et al	'65	739	Delinquent Youths	Shows CPI & BE best combined prediction of recidivism.
574-4320	Hindelang	'63	582	High School Boys	Shows relations of certain personality clusters to types of claimed delinquency.
474-0028	Jesness, C. F.	'73	3197	Delinquents & Control Boys & Girls	A study using CPI & MMPI to validate Jesness Inventory.
574-4704	Neithercutt	'68	201	Federal Parolees	A dissertation studying parole outcomes.
474-4003	Reckless et al	'57	300	6th Grade Boys	Uses CPI to delineate personality.
474-4004	Reckless et al	'56	126	"Good" Boys	Uses CPI to delineate personality.

TABLE 8
STUDIES USING THE CALIFORNIA PSYCHOLOGICAL INVENTORY

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
574-3856	Siegman	'62	79		Anonymous Admitted criminal behavior in college students.
474-3998	Stein et al	'70	996	High School Boys	Compared High & Lower So Scores on delinquency, not a test of true predictions.
474-4020	Stein et al	'66	560	High School boys	Used CPI to test ability to discriminate delinquents from non-delinquents.
79 574-4135	Stein, K.	'68	971	Prisoners	
574-4234	Stein et al	'71	346	Delinquent Boys	Shows 7 types of delinquents.
674-5011	Werner	'72	934	CYA Delinquents	Uses CPI to describe subjects.
574-4176	Whinery	'71	214	16-18 Year Old Traffic Offenders	Studied effect of sentencing.

CHAPTER V
REFERENCES

1. Lachar, D., "Accuracy and Generalizability of an Automated MMPI Interpretation System", Journal of Consulting and Clinical Psychology, 1974, 42, pp. 267-273.
2. Monachesi, E. D. and G. R. Hathaway, "The Personality of Delinquents," in Butcher, J. N. (Ed.) MMPI: Research Developments and Clinical Applications, New York: McGraw-Hill, 1969, pp. 207-219.
3. Flanagan, J. J. and G. R. Lewis, "First Prison Admissions with Juvenile Histories and Absolute First Offenders: Frequencies and MMPI Profiles," Journal of Clinical Psychology, 1974, 30, pp. 358-360.
4. Gregory, R. J., "Replicated Actuarial Correlates for Three MMPI Code Types in Juvenile Delinquency," Journal of Clinical Psychology, 1974, 30, pp. 390-394.
5. Cottle, W. C., "Card Versus Booklet Forms of the MMPI," Journal of Applied Psychology, 1950, 34, pp. 255-259.
6. Kincannon, J. S., "Prediction of the Standard MMPI Scale Scores from 71 Items: The Mini-Mult", Journal of Consulting and Clinical Psychology, 1968, 32, pp. 319-325.
7. Armentrout, J. A. and D. L. Rouzer, "Utility of the Mini-Mult with Delinquents", Journal of Consulting and Clinical Psychology, 1970, 34, p. 450.

8. Mlott, G. R., "The Mini-Mult and its Use with Adolescents," Journal of Clinical Psychology, 1973, 29, pp. 376-377.
9. Hobbs, J. R., "Scale Equivalence and Profile Similarity of the Mini-Mult and MMPI in an Outpatient Clinic", Journal of Clinical Psychology, 1974, 30, pp. 349-350.
10. Panton, J. H., "MMPI Profile Configurations Among Crime Classification Groups," Journal of Clinical Psychology, 1958, 14, pp. 305-308.
11. Christensen, L. and A. LeUnes, "Discriminating Criminal Types and Recidivism by Means of the MMPI," Journal of Clinical Psychology, 1974, 30, pp. 192-193.
12. Wattron, J. B., "A Prison Maladjustment Scale for the MMPI," Journal of Clinical Psychology, 1963, 19, pp. 109-110.
13. Mack, J. L., "The MMPI and Recidivism," Journal of Abnormal Psychology, 1964, 74, pp. 612-614.
14. Jaman, D. R., "The Behavior During the First Year in Prison, Report II, MMPI Scales and Behavior," Research Report No. 34, Research Division, Department of Corrections, State of California, 1969.
15. Whinery, L. H. Project on Predictive Sentencing of 16-18 Year Old Male Habitual Traffic Offenders, Municipal Criminal Court of Norman, Oklahoma, Sept., 1971.

16. Shupe, D. R. and P. F. Bramwell, "Prediction of Escape from MMPI Data," Journal of Clinical Psychology, 1963, 19, pp. 223-226.
17. Beall, H. G. and J. H. Panton, "Use of the Minnesota Multiphasic Personality Inventory as an Index to Escapism", Journal of Clinical Psychology, 1956, 12,
18. Jaman, Dorothy R., Patricia Coburn, Jackie Goddard and P. F. Mueller, Characteristics of Violent Prisoners, California Youth and Adults Corrections Agency, June, 1966.
19. Wenk, E. A., J. O. Robison and G. W. Smith, "Can Violence be Predicted?" Crime and Delinquency, 1972, (Oct.) pp. 393-402.
20. Megargee, E. J. and G. A. Mendelsohn, "A Cross-Validation of Twelve MMPI Indices of Hostility and Control," in Megargee, E. J., (Ed.) Research in Clinical Assessment, New York: Harper and Row, 1966, pp. 282-292.
21. Blackburn, R., "Dimensions of Hostility and Aggression in Abnormal Offenders," Journal of Consulting and Clinical Psychology, 1972, 38, pp. 20-26.
22. Deiker, T. E., "A Cross-Validation of MMPI Scales of Aggression on Male Criminal Criterion Groups," Journal of Consulting and Clinical Psychology, 1974, 42, pp. 196-202.
23. R. B. Cattell and H. W. Eber, Sixteen Personality Factor Questionnaire, Institute for Personality and Ability Testing.

24. Becker, N. C., "A Comparison of the Factor Structure and Other Properties of the 16PF and the Guilford-Martin Personality Inventories," Educational and Psychological Measures, 1961, 21, pp. 393-404.
25. Adcock, C. J., "A Review of the 16PF Questionnaire," in Buros, O. K. (Ed.), The Fifth Mental Measurements Yearbook, New Jersey: Gryphon Press, 1959, p. 112.
26. Cowden, J. E., W. M. Peterson and M. F. Cohen, "The 16 PF vs. the MCI in a Group Testing Program within a Correctional Setting," Journal of Clinical Psychology, 1968, 24, pp. 221-224.
27. Cowden, J. E., C. R. Schroeder and W. M. Peterson, "The CPI vs. the 16PF at a Reception Center for Boys," Journal of Clinical Psychology, 1971, 27, pp. 109-111.
28. Adcock, C.J., op. cit.
29. Hogan, Robert, "Review of the HSPQ", in Buros, O. K. (Ed.) The Seventh Mental Measurements Yearbook, p. 97, New Jersey: Gryphon Press, 1972.
30. Jackson, Douglas, "A Review of the HSPQ", in Buros, O. K. (Ed.) The Seventh Mental Measurements Yearbook, New Jersey: Gryphon Press, 1972, p. 97.
31. White, W. F. and T. L. Porter, "Multivariate Analysis of Attitudes and Personality Characteristics among 60 Youthful Offenders", Psychological Reports, 1970, 26, pp. 487-491.

32. McQuaid, John, "A Personality Profile of Delinquent Boys in Scottish Approved Schools," British Journal of Criminology, 1970.
33. Pierson, George R., John Moseley and Mark Olsen, "The Personality and Character Structure of the Delinquent: Some Social Psychological Implications," Journal of Genetic Psychology, 1967, 110, pp. 139-147.
34. Goodstein, Leonard D., and Schroeder, William J., "An empirically-derived managerial key for the California Psychological Inventory," Journal of Applied Psychology, 1963, 47, p. 42-45.
35. Hogan, Robert, "Development of an empathy scale," Journal of Consulting and Clinical Psychology, 1969, 33, p. 307-316.
36. Leventhal, Allan M., "An anxiety scale for the CPI," Journal of Clinical Psychology, 1966, 22, p. 459-461.
37. Kirk, Barbara A., Cummings, Roger W., and Hackett, Herbert R., "Personal and vocational characteristics of dental students," Personnel and Guidance Journal, 1963, 41, p. 522-527.
38. Cowden, J. E., C. R. Schroeder and W. M. Peterson, op. cit., p. 109-111.
39. Stein, K. B., A. Gough and T. R. Sarbins, "The Dimensionality of the CPI Socialization Scale and an Empirically - Derived Typology among Delinquent and Non-Delinquent Boys," Multivariate Behavioral Research, 1966, 1, pp. 197-208.

40. Gough, H. A., "Theory and Measurement of Socialization," Journal of Consulting Psychologists, 1960, 24, pp. 23-30.
41. Stein, K. B., A. C. Vadum and T. R. Sarbin, "Socialization and Delinquency: A Study of False Negatives and False Positives in Prediction," Psychology Record, 1970, 20, pp. 353-364.
42. Gough, H. A., "Systematic Validation of a Test for Delinquency", (Abstract), American Psychologist, 1954, 9, p. 381.
43. Hindelang, M. J., "The Relationship of Self-Reported Delinquency to Scales of the CPI and MMPI," Journal of Criminal Law, Criminology and Police Science, 1972, 63, pp. 75-81.
44. Argyle, M., "A New Approach to the Classification of Delinquents with Implications for Treatment," Inquiries Concerning Treatment for Delinquents, pp. 15-26, California State Board of Corrections, 1961.
45. Neithercutt, M., "Predicting Outcomes of Federal Parolees," Dissertation Abstracts, 1968, 69, p. 3538.
46. The Community Treatment Project After Five Years, Department of Youth Authority of State of California, California Youth Authority, 1966.
47. Waldo, G. P. and S. Dinitz, "Personality Attributes of the Criminal: An Analysis of Research Studies," 1950-65, J. Res'ch in Crime and Delinquency, 1967, 4, pp. 185-202.

48. Gough, H. G., E. A. Wenk, and V. V. Rozyngo, "Parole Outcome as Predicted rom the CPI, the MMPI and a Base Expectancy Table," Journal of Abnormal Psychology, 1965, 70, pp. 432-441.
49. Jesness, C. F., M. Bohnstedt, M. J. Moloff and R. Wedge, Sequential I-Level Classification, American Justice Institute, California Department of Youth Authority, 1973.
50. Jesness, C. F., The Jesness Inventory, Consulting Psychologists Press, 1966, Palo Alto.
51. Ibid
52. Cowden, J. E., W. M. Peterson and A. R. Pacht, The MCI vs. the Jesness Inventory as a Screening and Classification Instrument at a Juvenile Correctional Institution.
53. Griffiths, K. S. (Ed.), A Review of Accumulated Research in the California Youth Authority, California Youth Authority, 1974.
54. Weintraub, S. A., "A Review of the Jesness Inventory", in Buros, O. K. (Ed.) The Seventh Mental Measurements Yearbook, New Jersey: The Gryphon Press, 1972.
55. Fraas, L. A. and R. L. Price, "The Jesness Inventory as a Prediction of AWOL Recidivism," Psychology Reports, 1972, 31, pp. 741-742.
56. Kelly, F. J. and D. V. Baer, "Jesness Inventory and Self Concept Measures for Delinquents Before and After Participation in Outward Bound," Psychologists Reports, 1969, 25, pp. 719-724.

57. Warren, M. Q., T. B. Palmer, V. V. Neto and J. K. Turner, "Community Treatment Project: An Evaluation of Community Treatment for Delinquents," Fifth Progress Report, Community Treatment Project #7, 1966.
58. Butler, E. W. and S. N. Adams, "Typologies of Delinquent Girls: Some Alternative Approaches," Social Forces, 1966, 44, pp. 401-407.
59. Jesness, C. F., "The Preston Typology Study: An Experiment with Differential Treatment in an Institution," Journal of Crime and Delinquency, 1971, 8, pp. 38-52.
60. Warren, M. Q., "The case for Differential Treatment of Delinquents," Annals American Academy of Political and Social Science, 1969, 381, pp. 47-59.
61. Richards, W. C., C. G. Mates and L. Whitten, "Personality Traits and Attitudes of Adolescent Girls With Behavior Disorders," Correctional Psychiatry and Journal of Social Corrections, 1969, 5(2), pp. 34-44.
62. Fitts, William H., and Hamner, William T., The Self Concept and Delinquency, Monograph I, Nashville, Tennessee, Counselor Recordings and Tests, 1969.

CHAPTER VI
INTELLECTUAL MEASURES

The Wechsler-Bellevue Intelligence Scale was originally prepared to satisfy an existing need "for an individual examination devised primarily for and adapted specifically to the measurement and appraisal of adult intelligence". In addition, it has established itself over the years as a basic psychological diagnostic instrument. It has been revised and renamed the Wechsler Adult Intelligence Scale (WAIS), though the changes in format are minimal.⁽¹⁾ It has also been extended downward to form the Wechsler Intelligence Scale for Children (WISC).⁽²⁾

The test is standardized on a nationwide sample of 1700 adults, prorated according to the 1950 U. S. Census, including a proportionate sample of the non-white population. The WAIS consists of eleven subscales:

Verbal

Information
Comprehension
Arithmetic
Similarities
Digit Span
Vocabulary

Performance Tests

Digit Symbol
Picture Completion
Block Design
Picture Arrangement
Object Assembly

The reliability coefficient for the Verbal I.Q. is reported to be .96; for the Performance I.W., .93 and .94; and .97 for the Full Scale I.Q. However, reliability coefficients for the sub-scales, although nearly identical over the age range used, varied from .60 to .96 for different scales. Obviously, the significance of scores on scales with low reliability must be evaluated cautiously. Special caution is needed in comparing differences between or within profiles.

Validity of the scale has been determined by the method of standardization and statistical treatment of the items and by correlation with the Stanford-Binet, for example, as well as with many other tests and measures of intellectual performance or ability. By general consensus, and years of use by thousands of examiners in widely varied settings, it must be conceded to be the best single measure of intelligence available.

But almost from the date of its first publication there has been an attitude of regarding intelligence as a personality variable distinct from other personality variables. Among other areas of interest, there has been a significant effort to use WAIS I.Q.s as predictive of other specific performance. In terms of academic success this effort has been highly successful in generalized terms, but considerably less so successful in attempting to predict success in specific curricular activities.

Wechsler has stated that the most significant single feature of the sociopath's test profile is that his Performance I.Q. tends to be consistently higher than the Verbal I.Q.⁽³⁾ This has been

confirmed by many investigators for mean scores of groups, but not for individuals. Some have not even been able to confirm this relationship for groups.⁽⁴⁾

Henning and Levy, in a very thorough study, show that there are many possible sources for the discrepancies, such as cultural factors, environmental factors, reading disability, mental retardation, and subtle differences between the WISC and WAIS scales themselves.⁽⁵⁾

A review of the results obtained from these tests and many others, leads us to conclude we cannot identify "sociopaths" or delinquents or schizophrenics solely by the use of intelligence tests, although the use of such tests may be of considerable value in predicting the relationship between behavior and personality.⁽⁶⁾ Studies using the WAIS and the WISC are shown on Tables 9 and 10.

TABLE 9
STUDIES USING THE WECHSLER ADULT INTELLIGENCE SCALE

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
474-3400	Blackburn, R.	'72	165	Abnormal Offenders	Uses WAIS to get I.Q. as part of Career Study of "Abnormal Offender"
474-3363	Climent, C., et al	'73	95	Women Prisoners	Got WAIS on violent female offenders as part of other data.
473-3367	Deiker, T.	'73	243	Murderers	Gives murderers scores, but no comparison group.
474-3188	Henning, J. & Levy, R.	'67	2361	Delinquents	Part of evaluation of WISC and WAIS for delinquents.
574-4136	Ross	'72	84	Female Addicts	The WAIS was used to validate vocabulary, intelligence tests.
574-4156	Svenson & Grimes		45	Sex Offenders	Seems descriptive of schools and social history.
474-3851	Van Vorst	'43	no data	Psychopatic	Data not given.
574-4635	Wagner	'64	60	Aggressive Delinquent	Tested different responses between 2 groups - Assaultive and non-assaultive show them significantly different.
474-3162	Young, F.	'56	68	Delinquents	Simple description of population as given by TAT responses.

TABLE 10
STUDIES USING THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN
(WISC)

REFERENCE	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
474-3188	Henning, J. & Levy, R.	'67	2361	Delinquents	Showed I.Q. differences between Black & White Delinquents
574-4145	Hurwitz, I. et al	'72	45	Delinquents compared to normals & boys with learning problem	WISC used to control on I.Q., was not independent variable.

CHAPTER VI
REFERENCES

1. Wechsler, D., The Measurement and Appraisal of Adult Intelligence, Baltimore: Williams & Wilkins Co., 1958.
2. Freides, David, "A review of the Wechsler Intelligence Scale for Children," in Buros, O. K., (Ed.) The Seventh Mental Measurements Yearbook, New Jersey: Gryphon Press, 1972, p. 431.
3. Wechsler, op. cit.
4. Matarazzo, J. D., Wechsler's Measurement and Appraisal of Adult Intelligence, Baltimore: Williams & Wilkins, 1972.
5. Henning, J. J. and R. H. Levy, "Verbal Performance I.Q. Differences of White and Negro Delinquents on the WISC and WAIS", Journal of Clinical Psychology, 1967, 23, pp. 164-165.
6. Burstein, R. G., "A Review of the Weschsler Adult Intelligence Scale," in Buros, O. K., (Ed.) The Seventh Mental Measurements Yearbook, New Jersey: Gryphon Press, 1972, p. 429.

CHAPTER VII
MEASURES OF EDUCATIONAL LEVEL

The Otis-Lennon Mental Ability Test has been standardized on a sample chosen to represent the country's educational system, not the population at large. It reflects the highest professional standards in construction, norming, reliability and validity. It is not intended as a measure of innate learning potential, and users are warned against interpreting results of individuals who lack normal backgrounds and motivation. Unfortunately, it has sometimes been used for purposes for which it was not intended.⁽¹⁾

The test is already widely used in institutional classification. It will provide deviation I.Q.s as well as provide norms for various grade levels and grade equivalents for such specific subject matter areas as reading, arithmetic, social studies and so on. The deviation I.Q.s are normalized standard scores with a mean of 100 and a standard deviation of 16 points. Thus, the scores are comparable at all age or educational levels. Grade percentile ranks and stanines have also been developed and permit additional interpretation.

The Stanford Achievement Test has been published in various editions since 1922. The various test batteries cover the range from grade 1.5 through grade 12. It is available in Braille or large type. The various forms and editions measure every standard high school subject, and some less common ones. Reliabilities

are generally in the .80's or .90's, though some of the subject matter tests have reliabilities in the .60's. Validity data are not given. Norming is excellent, and chosen from varied geographic areas. Norms for the high school test are based on a sample of 22,699. Testing time for the entire high school battery is 350 minutes, and it is suggested the test be given in six sessions.⁽²⁾

The California Achievement Test (CAT) has been published in various editions since 1934. The most recent edition is the 1970. This edition is in two forms, five levels covering from grades 1.5 to grade 12. Scores are reported for reading, mathematics, language measured in nine subtests at the 9 to 12 grade level. Each of the three major divisions is available in a separate booklet. Each section requires approximately one hour to administer. Depending upon the kind of answer sheet used, the tests may be scored by hand or machine.

Norms for the 1970 Edition are based on over 203,684 students in rural, suburban, and urban schools in all parts of the country. The test has been well prepared and well standardized. The Manual presents a wealth of data and extensive norms, far more than most users will require.

Reliability, as determined by the Kuder-Richardson 20 formula, is in the high .80's or .90's for all tests at all grade levels except for languages. In this test reliability falls to the .60's or .70's. Validity is based primarily on method of test construction and item solution, too technical and detailed to permit discussion here. However, two technical manuals present the detailed information for those interested. Table 11 shows an offender study using the CAT.

The Wide Range Achievement Test is a measure of reading, spelling and arithmetic, ordinarily administered individually, but with provision for group administration of some parts. It is intended as an adjunct to individual clinical evaluation, and is impractical for general school use. Reliability and validity are highly satisfactory and the results may be used diagnostically by an experienced and skilled diagnostician.^(3,4)

The Peabody Picture Vocabulary Test (PPVT) consists of 150 numbered plates, each with 4 pictures. As the test is intended for use with illiterates, the subject does no reading or writing, but simply points to the picture that matches the word read aloud by the examiner. The examination requires only 10 to 15 minutes. Raw scores may be expressed as mental ages, standard scores (I.Q.) or percentile equivalents. Reliability coefficients range from .67 at the 6 year level to .84 at 17 and 18 year old levels.⁽⁵⁾

Congruent and concurrent validities are reported, ranging from .30 to .84 depending on the age of the individual and the specific test used with which to evaluate PPVT. Studies of predictive validity are too few to be of value at present. However, the test appears to be a highly useful instrument for evaluating intellectual level with an illiterate population, observing the usual cautions regarding age, social or cultural level, motivation and so on.

A. Career Planning Measures

It would appear to be almost self-evident that an individual reaching adult status unprepared to earn a living is more likely to be a candidate for illegal behavior than one who can fit usefully and constructively into society. Other factors are involved, of course; not all unskilled workers are criminals - not all criminals are unskilled. But one who can do something useful is more likely to be able to be employed and therefore more likely to be paroled. If he is thus able to support himself, he is presumably less likely to resort to crime. The proportion of those unprepared for any specific occupation is higher in the criminal population than in the non-criminal population. This seems to be at least a significant part of the basis for providing occupational training while the individual is confined.

Once the decision is made to provide occupational training, two questions arise immediately:

1. What kind of training is to be provided?
2. How is the offender to be assigned to what specific training?

The training opportunities provided will be determined by many factors, practical, political, and economic among others. This is not the time or the place to discuss what training is to be offered in any institutional or community program, though it may be noted that apparently most such training is manual and/or

mechanical in nature. To some extent this stems from the traditional viewpoint of imprisonment as "labor", in part from the uneducated and illiterate population which constitutes the bulk of the inmates, in part from the needs of the institution and in part from the restrictions imposed by society.

The General Aptitude Test Battery (GATB) is probably the best multiple aptitude test available for evaluating career probabilities. The extensive use of the test with traditional "blue collar" jobs has made its occupational patterns highly useful to counselors and placement officers. However, the lack of adequate research in the "white collar" occupations limits its usefulness. It has a great deal of validity data of the traditional sort, though it does not provide adequate data to show that an individual with a high score can learn to do a job better than an individual with a low score. Thus, it must be interpreted as a measure of current status rather than as ability to learn.⁽⁶⁾

Many of the hard core unemployed, and many offenders are illiterate or semi-illiterate, or come from a cultural background different from the traditional white, native American, middle-class on which so many widely used tests have been standardized. To meet the needs of these disadvantaged individuals the United States Training and Employment Service has developed the Non-reading Aptitude Test Battery (NATB). The test is used in conjunction with the GATB, and interpretation is based on the GATB norms and occupational patterns. The test is too new to have independent reports of its value in print, but is apparently adequately standardized.⁽⁷⁾

TABLE 11
STUDIES USING THE CALIFORNIA ACHIEVEMENT TEST
(CAT)

REFERENCE NO.	AUTHORS	DATE	SIZE	SAMPLE POPULATION	CONCLUSION
574-4135	Stein, K.	'68	971	California prisoners	Showed groups differing on MAIPS also differed on other scales, including CAT.

CHAPTER VII
REFERENCES

1. Milholland, J. E., "A Review of the Otis-Lennon Mental Ability Test," in Buros, O. K. (Ed.) The Seventh Mental Measurements Yearbook, New Jersey: Gryphon Press, 1972, p. 429.
2. Adams, Georgia S., "A Review of the Stanford Achievement Test", in Buros, O.K. (Ed.) The Seventh Mental Measurements Yearbook, New Jersey: Gryphon Press, 1972, p. 27.
3. Merwin, J. C., "A Review of the Wide Range Achievement Test, Revised," in Buros, O. K. (Ed.) The Seventh Mental Measurements Yearbook, New Jersey: Gryphon Press, 1972, p. 36.
4. Jastak, J. F. and S. R. Jastak, The Wide Range Achievement Test, Wilmington, Delaware: Guidance Associates of Delaware, 1965.
5. Lyman, H. G., "A Review of the Peabody Picture Vocabulary Test," in Buros, O. K., (Ed.) The Sixth Mental Measurement Yearbook, New Jersey: Gryphon Press, 1965, p. 530.
6. Weiss, D. J., "A Review of the General Aptitude Test Battery," in Buros, O. K. (Ed.) The Seventh Mental Measurements Yearbook, New Jersey: Gryphon Press, 1972, p.676.
7. Dragee, Robert C., William Showler, Stephen Bemis and John Hawk, "Development of a nonreading edition of the GATB", Measurement and Evaluation Guidance, 1970, 3, pp. 45-53.

CHAPTER VIII
EXPRESSION OF INTEREST AND VALUES

The Kuder General Interest Survey (KGIS) is a revision and downward extension of several earlier vocational preference inventories. Test users who have the Kuder Preference Record Vocational (KPRV) will find little difference between the two. However, the KGIS is simpler in language, and now has a 6th grade reading level. Both inventories report interests in the same ten occupational scales: outdoor, mechanical, computational, scientific, persuasive, artistic, literary, musical, social service, and clerical. There is also a verification scale which is intended to indicate whether or not the scale has been properly completed. This interest inventory is the one most widely used in contemporary correctional institutions.

The Manual for this inventory presents lists of occupations classified according to major interest patterns. The scores are reported in a profile form so that major interests are easily and quickly identified.

Because the KPRV is so easy to administer, score and interpret, it has been widely accepted, generously used, and often misinterpreted. It seems probable that KGIS will be received in the same way. The author, in the test manual, cautions the test user, but our own experience indicates that many users of this survey will not read the manual, and if they do, will not employ the caveats in their actual practice.

With this caution formally stated, it is nevertheless necessary to say that the KGIS is the result of extended research and effort at a very high level, and that the manual contains a wealth of information for those who will use it. (1,2)

There are a number of other interest inventories available which for special reasons or special purposes, may be desirable. The most widely used is the Strong Vocational Interest Inventory. This has not, however, been widely used in correctional institutions. The reasons for its less frequent use are probably its greater complexity, its greater difficulty both in administration and scoring, its greater cost, and the fact that it is focused primarily on business and professional occupations.

CHAPTER VIII
REFERENCES

1. Kirk, Barbara A., "Review" in Buros, O. K. (Ed.) The Seventh Mental Measurements Yearbook, New Jersey: Gryphon Press, 1972, p. 1024.
2. Lohnes, Paul R., "Review" in Buros, O. K. (Ed.) The Seventh Mental Measurements Yearbook, New Jersey: Gryphon Press, 1972, p. 1024.

CHAPTER IX
ASSESSMENT OF SOCIAL ADJUSTMENT

The Quay Battery consists of three tests or scales - the Behavior Problem Checklist (BPC), the Checklist for the Analysis of Life History Data (CALH), and the Personal Opinion Study (POS).⁽¹⁾ These are designed through factor analysis to measure four deviant personality types which Quay has been refining and developing over the last 15 years. These are called:

- BC1 - Inadequate - Immature
- BC2 - Neurotic - Disturbed
- BC3 - Unsocialized - Psychopathic
- BC4 - Socialized - Subcultural
- and a fifth, tentative one,
- BC5 - Which is a subgroup of 1 and 4

The BPC is a three point rating scale with 55 items describing deviant behavior. It is designed for use by correctional personnel, professionals, parents, teachers, anybody who knows the person well enough to rate him. Typical dimensions on which the person is rated are: easily flustered, hypersensitivity, anxiety, short attention span, irresponsible, disruptive. It has been purified and given internal consistency through extensive factor analysis. It measures BC1, 2 and 3.

The CALH is a checklist designed to be used by the caseworker familiar with the person's life history. There are 36 items to be checked such as: engages in furtive stealing, incompetent, habitual

truant, seclusive, submissive, assaultive and has bad companions. This adds a fourth category, BC4, Socialized-Subcultural. Quay also suggests a fifth tentative factor, BC5, which is a subtype of 1 and 4. The checklist has been submitted to the same rigorous factor analysis.

POS consists of 100 items to be marked True or False by the subject himself. In this it resembles the standard inventory like the MMPI or the CPI. Typical items are: I feel tired a good bit of the time; In this world you are a fool if you trust other people; It is important to have enough friends and social life.

The use of standard and T scores makes scores additive across the three scales to give a composite score and to furnish cross comparisons among the subtypes. The use of different raters and the subject's own responses adds a further type of practical reliability and validity to the battery.

Reliability is good, internal, repeat, and inter-rater are all reported and at good levels; validity is excellent. Careful factor analysis in tailoring the battery gives it internal consistency. The selection of items furnishes face validity. Concurrent validity is demonstrated by its agreement with outside criteria such as the Gough Socialization Scale, etc. Construct validity is furnished by the careful theoretical rationale in which it has been developed.

An excellent job of scale development, as test per se, it is probably the best in the lot designed for classification purposes. The only question is whether or not the categories proposed offer

the best basis for a classificatory system designed to meet correctional needs. In turn this rests on what alternative treatments are available in the system and the presence of differentially adapted or trained personnel to carry them out. All these are questions for the future.⁽²⁾

The Environmental Deprivation Scale (EDS) is an interview guide that covers 16 items in three broad occupational, institutional and interpersonal areas. Each item is scored "0" for positive input and "1" for negative input. Total score is the sum. The 16 items are:

- | | |
|------------------------|-------------------------|
| 1. Employment | 9. Church |
| 2. Income | 10. Other organizations |
| 3. Debts | 11. Friends |
| 4. Job participation | 12. Relatives |
| 5. Job status | 13. Parents |
| 6. Hobbies & avocation | 14. Wife |
| 7. Education | 15. Children |
| 8. Residence | 16. Fear |

In an interesting 1971 follow-up study of 128 offenders after release from prison, DeVine et al consistently refer to the "accuracy of prediction" as an expression of validity.⁽³⁾ However, it should be noted that they are reporting concurrent validity rather than prediction. (See p.27 Chapter III for further discussion).

Validated against offenses subsequent to release, as measured by the Law Encounter Severity Scale (LESS), there appears to be a significant discriminating relationship the lower the EDS, the more

likely is LESS to be low as well. They also report that the greater the deficit in the three areas, the more severe was the law encounter. This group did not find differences in institutional treatment reflected in changes in the EDS scores. They did find that low EDS scores tend to remain low over a period of time, but high EDS scores tend to become higher.

The differentiation this group reports in their study of the EDS and LESS should encourage a predictive study of parole success. If future studies confirm the relationship, the EDS will become a highly valuable scale to use in conjunction with an experience table.

Of significance for the value of career planning and training, also, is their conclusion that the highly significant validities of the occupational cluster indicates a need for "selection and training in a vocational area that is highly reinforcing to the individual".⁽⁴⁾ (See p. 49 , Chapter V, for further discussion).

The original standardization of the EDS was validated and cross-validated on samples of 173 and 142 subjects. Again, although the authors consistently refer to "prediction", they do not report predictive data. The concurrent nature of the validity studies, although impressive, may be contaminated by subjective judgement of the interviewers.

This scale is designed to be accompanied by the Maladaptive Behavior Record (MBR) as a measure of response outputs. In a structured interview, the MBR includes sixteen topics covering addiction, interpersonal relations, economics, psychological

adjustment, etc. It was validated on 216 male felony offenders and is subject to the same criticism of reliance on subjective judgement, although reliability on retest is fairly good.⁽⁵⁾

The final measure in what essentially constitutes a battery is the Weekly Activity Record (WAR). It was validated on 117 prison releasees, with data gathered in an interview. It has been validated in terms of graduated means against five groups of severity on the Law Encounter Severity Scale. Trend data again emerge but no significant data on differences in means.

Reliability data are unclear. The authors report that they have replicated a previous study, but do not present data showing the amount of agreement or difference. More serious, however, is the possible contamination of the data due to the time of the sampling, relations to release from prison, and for some, the proximity of interview to recidivism. As each subject was interviewed three times, data on the agreement-or lack of it-of successive WAR scores would be indicative of reliable consistency of behavior.⁽⁶⁾

Accompanying these is the Law Encounter Severity Scale (LESS), a 38 point ordinal scale, which forms a law encounter continuum. The items are ranked in order of severity, even including those that did not involve a sentence or fine. Five major groups of items emerged from ranking by three judges. There was agreement in ranking of 90-95% of the items, and complete agreement in the five groups of items. LESS is thus a scale for all types of law encounters, and may be used with adults, juveniles, parolees, and others. It was developed to become the criterion by which to validate EDS, MBR, and WAR.⁽⁷⁾

General criticisms are weak validity data and reliance on subjective judgement although the dimensions are suggestive.

CHAPTER IX
REFERENCES

1. Quay, H. C. and L. B. Parsons, The Differential Behavioral Classification of the Juvenile Offender, Second Edition, Bureau of Prisons, 1971.
2. Bureau of Prisons, U. S. Department of Justice, Washington, D. C. 20537, September, 1970, Differential Treatment ... a way to begin.
3. DeVine, M. D, W. O. Jenkins, A. D. Witherspoon, E. K. deValera, J. B. Muller and J. M. McKee, The Environmental Deprivation Scale (EDS): The Role of Environmental Factors in the Analysis and Prediction of Criminal Behavior and Recidivism, Montgomery Ala: Rehabilitation Research Foundation, 1974.
5. Barton, W. and W. O. Jenkins, The Maladaptive Behavior Record (MBR): A Scale for the Analysis and Prediction of Community Adjustment and Recidivism of Offenders, Experimental Manpower Laboratory for Corrections, Montgomery, Alabama, January, 1973.
6. Jenkins, W. O., J. B. Muller, M. D. DeVine, E. K. deValera, A. D. Witherspoon and J. M. McKee, The Weekly Activity Record: A Measure of Time Allocation in the Analysis and Prediction of Criminal Behavior and Recidivism, Experimental Manpower Laboratory for Corrections, Montgomery, Alabama, April, 1974.

7. Witherspoon, A. D., E. K. deValera and W. O. Jenkins, The Law Encounter Severity Scale (LESS): A Criterion for Criminal Behavior and Recidivism, Experimental Manpower Laboratory for Corrections, Montgomery, Ala. August, 1973.
8. Jenkins, W. and W. Sanford, A Manual for the Use of the Environmental Deprivation Scale in Corrections: The Prediction of Criminal Behavior, Experimental Manpower Laboratory for Corrections, Montgomery, Alabama, 1972.

CHAPTER X
A SUGGESTED PROGRAM OF TEST
AND ASSESSMENT FOR ADULT OFFENDERS

To plan a testing program to be used in assessing or evaluating adult offenders, we must begin with a simple statement of the goal to be achieved. This might be in such form as: "To obtain the maximum useful information with the least effort, least time, and least cost." We will be concerned with the materials used both initially and subsequently. The amount of effort involved and the time required for various steps in the procedure will concern us as well. The kind of skill or training required will have a significant bearing on the desirability of the proposed test.

It also seems evident that the greater the homogeneity of testing at the federal, state and local level, the greater will be the clarity of communication. The more parsimonious the test battery, provided it assesses the needed areas, the easier it is to get a clear picture of the test results. A small battery is much more likely to be uniformly adopted, despite the fact it may provide minimal information. It is, of course, obvious that whatever the battery, it is only as useful as the expertise available to interpret. It is, therefore, assumed that skilled persons, experienced in test administration and interpretation are available for decision making affecting the life of the offender.

The nature of the results obtained will be of primary concern. We will need to ask if the instrument is valid and reliable, if the results obtained are in a form that can be understood by all who need to evaluate them, and in such a form that they can be applied to the particular purpose for which the tests were administered. Each of the questions must be answered individually, but all factors must be considered in the development of the battery of instruments to be used.

In the following discussion we have attempted to consider variations in the amount of skill required for administration, the amount of time required to administer, the time, effort and skill needed to understand the results, and the relative costs of these steps. We have tried to develop a basic battery which would be administered to all individuals entering the correctional program; a diagnostic battery which would be used to further explore those who appeared to present some problems; a supplementary battery for those who, while not presenting problems, need further study to help understand their capacities and suitability for special programs while in prison, and a suggested research battery. We feel that if these suggestions are adopted by a number of centers, and the results coordinated, progress in classification and treatment can be greatly increased.

Before attempting to select the test battery or batteries, however, we must first define the purpose of the testing, and the kind of information we wish to obtain. As our first step is the grouping of offenders in order to plan their handling and treatment while in the facility, there appears to be five basic areas of investigation:

- a. Base expectancy or probability of recidivism
- b. Educational skills and background
- c. Intelligence, or learning ability
- d. Personality factors and adjustment
- e. Occupational interest, aptitude and ability

A. Base Expectancy Rates Determined by Prior Career

Currently the California Tables seem to be doing an adequate job of determining in general the likelihood that certain types of adult male offenders will relapse. There are, however, some cautions that need to be made.

The B-E tables of California and all similar tables (e.g., the Illinois Experience tables) have been developed for men only. The lack of expectancy tables for women is a major gap in classification efforts and such tables should be developed.

It has been well established that B-E or experience tables change over time. As the types of persons sent to prison change, and as the external environment to which they are released changes, the predictive ability of the tables change. The behavior of persons released on parole should be constantly fed back to the system so that these tables can be kept up to date.

Although such tables are reasonably well researched for parole purposes, there is little data for similar tables to predict the expected results from diversion, probation, and institutional performance. It is therefore imperative that they be quickly developed to help the courts, probation staff and institutional workers make reasonable judgements about their clients.

Once such tables are developed a similar feedback system should be developed to keep these tables up-to-date. The development of such tables is dependent totally on having reasonable volumes of data available to a research staff continuously working on experience tables development.

Finally it should be noted that it is important that populations of "normals" be included. This would give some knowledge of the incidence of criminal behavior in "normal" populations and would indicate "normal" crime levels. Insurance life tables start with total populations, not with populations with various illnesses.

B. Education

An efficient use of the testing program depends upon the ability of the offender to respond meaningfully to the instruments used. Efficient use of time requires reading skill. Consequently, our first test will be the Reading subtest of the California Achievement Test series. If the reading level is below the ninth grade, we will shift from the Basic Battery to the Diagnostic.

If the reading level is below the ninth grade level, the Wide Range Achievement Test would be administered in order to evaluate the educational skills of basic reading, spelling, writing and arithmetic.

The WAIS, an individually administered test would be substituted for the OLMAT. The WAIS and WRAT scores would be evaluated to determine the nature of the problem. This might be normal or bright intelligence but poor educational background,

or a generally quite retarded individual with poor learning potential, or a bright but emotionally disturbed person, or some other combination.

Further testing would be directly influenced by the results of the WRAT and WAIS. For example, non-verbal tests of personality, occupational aptitude and interest would be immediately substituted for the usual Basic Battery tests.

If the individual is capable of learning, but educationally deficient, and if the institution has resources for instruction, a plan for educational improvement might be established based on the results of the WAIS, CAT and WRAT.

C. Intelligence

Our next requirement is to determine whether or not the individual is capable of responding to the situations and stimuli to which he will be subjected. We have already seen that I.Q. as such is not a determiner of behavior, so we are not interested in exact and precise measurement. At the same time, we want our approximation to be reasonably accurate, and a valid predictor of his ability to learn and benefit from treatment. Because the learning ability of prisoners will range over nearly the entire spectrum of ability, we will think in terms of possible educational and occupational training which is available while in custody, as well as his possible educational and occupational opportunities after release.

The OLMAT is recommended as the test to be included in the Basic Battery to measure learning ability. As has been previously indicated, this is a highly valid and reliable measure of readiness to benefit from instruction and "ability to deal with the abstract manipulation of verbal, numerical and symbol systems of our culture."

If performance on the OLMAT is inferior, and especially if he appears to have language difficulty, we recommend the administration of the Revised Beta. When intellectual retardation is suspected, we suggest the use of the WAIS.

D. Personality and Character

The Basic test in this category would be the MMPI. We have previously dealt with this inventory in detail. It appears to be the most suitable instrument for general administration in terms of time, cost, ease of administration and interpretation, and particularly because of the great body of literature available with which to compare local results. It can be administered to non-reading individuals, and is available in foreign language versions including Spanish.

Wherever the budget of time, money and personnel permits, we would urge the inclusion of the 16PF as a supplementary test. Research has shown this test to be highly useful in certain situations. The items are derived from totally different sources, and the interpretation of results is quite different, so that the time spent would enrich, rather than duplicate, the interpretation obtained from the MMPI.

When the MMPI indicates serious maladjustment, individual evaluation will be necessary. Each clinician will, of course, select for his use those instruments and procedures with which he feels most comfortable, in terms of his own experience and training, and his preliminary evaluation of the nature of the problem he is to study. However, in order to standardize some basic elements of the appraisal, so that the exchange of results can be facilitated, we recommend that the minimum Diagnostic Battery include the Bender-Gestalt and the House-Tree-Person.

We have already noted the disagreements about scoring methods and meaning of these tests. They are, nevertheless, easy to give, relatively easy to score and interpret, and the results can be readily exchanged between researchers. Both yield objective results that can be related to a variety of behavior or other test material.

E. Occupational

Many individuals come into the correctional system with few, if any, marketable job skills. Thus, in diversion or probation the offender can be required to get vocational training. Institutions can provide such training. Before pressing offenders into vocational programs, however, it is necessary to know the probability of successful completion. One would not, for example, put a person who could never learn to draw a straight line into drafting.

At the present time, the General Aptitude Test Battery is widely used by the U. S. Training and Employment Service in its guidance for employment. This test can provide some information about what types of blue collar employment training are appropriate for different clients. It should, however, be realized that this test is not really helpful in selection for academic training, nor does it measure aptitude, but rather, it measures current skills.

The resources of the institution will, of course, affect the specific plan, but no other test or test battery appears to yield so much information for the same investment in time and money.

Because interest and motivation are so important a part of occupational success, we recommend the use of the Kuder Preference Record as a supplemental test. With interest and ability measured, the counselor and the inmate can agree on the most desirable and practical plan for occupational preparation.

When the GATB is impractical due to deficiency in educational skills, we recommend that the NATB be substituted.

F. Some Recommendations on the Classification of Minors

It is just as imperative that there be homogeneity in the testing for classification of minors as for adults. It unfortunately happens that there are serious deficiencies in what is available for minors.

CONTINUED

3 OF 4

First, there is no instrument for minors that is as well established and as widely applicable as the California Base Expectancy tables for adults. The California Youth Authority has developed Base Expectancies experience tables for its wards, but as was noted previously, this has not been proven generally applicable. For this reason, the very first task is the expansion of this or a similar table from constant feedback so that it is more broadly applicable than to California Youths alone. This is required before further serious development can be expected.

At the level of intelligence testing for program assignment, it is suggested that the Otis-Lennon Mental Ability Test be used, to be supplemented by the Peabody Picture Vocabulary Test (PPVT) for illiterates and all marginal cases. This test can reach from the age of approximately three years to the age of 18. Again, as with adults, these intelligence-education tests should be given prior to all other group tests, so individuals who are as yet unable to read at adequate levels will not be given inappropriate tests.

The second test recommended for adults was the MMPI. It is strongly recommended that it be used also with all juvenile offenders, 16 or over with an adequate reading level. For juveniles under 16 there seems to be no test as strongly established. The MMPI runs into problems of reading and vocabulary for children under 16 and for persons educationally limited or from a non-English speaking background. The examiner should see that the appropriate non-English version be used if one is available. It

seems necessary to urge that a standard screening device for persons 8 to 16, or for those with a reading level of from third to eight grade which can be group administered, be standardized.

A word must be said here about the Jesness Inventory, the I-Level Interview Schedule and the Quay battery. The tests are now of relatively long standing. Unfortunately, they have not been validated extensively in a general youth population. As a result their ability to predict behavior in an unselected adolescent population is not well-established. It is therefore strongly recommended that these three instruments be tested widely by researchers and clinicians who have not been identified with the development of the tests. What they purport to do is important. It must be clearly established at what levels the tests actually function. Hence we strongly recommend that the Jesness Inventory and the Quay battery be given on a routine basis to a predetermined fraction of all admissions or re-admissions. These facts should be analyzed carefully to determine their relative strengths with the hope that one or the other, or a new test, a composite of both, will provide a truly useful instrument on a national basis. The I-Level Interview Schedule should be given on a sample basis and its reliability and ability to discriminate between types of offenders and between offenders and non-offenders established on a nationwide basis.

For the assignment to learning programs by diagnosis of learning levels it is suggested that the California Achievement Test be universally adopted. The CAT is specifically recommended

for two reasons, first it seems to be already the most widely used in correctional circles and hence will require the least change. Secondly, its norms are at least as well established as any similar test in the field.

It should be clear that instrumentation for the classification of the youthful offender is by no means as well developed as for the adult. It is clearly imperative that common, nationwide efforts be taken to develop appropriate base expectancy tables and personality tests that will aid the correctional practitioner in understanding and developing programs for his clients. This implies clearly that there must be significant data collection on pre-test behavior, test results and post-test criminal and non-criminal behavior. Without such data, a truly meaningful and scientifically based system can never be developed.

G. Additional Comments

This discussion of tests has been focused almost entirely with the individual offender who has been committed to an institution. It is our hope, however, that the life history data required for the base expectancy tables be gathered prior to sentencing and reported centrally. We hope that all additional information including prison history and subsequent arrests, will be in like manner centrally gathered and analyzed. Thus continuous improvement of prediction can be made.

Not all offenders are arrested, not all those arrested are tried, not all those tried are convicted, not all those convicted

are sentenced, and not all those sentenced reach a prison. Decisions for direction are made at each step, by someone, on the basis of some kind of information. We can assume that each of these successive groups of people differ in some significant way from incarcerated offenders, but we cannot know how or to what degree they differ until a planned assessment has been made of sample populations at each of these steps.

Base expectancy tables are related to success on parole. Would the addition of the data from the Basic Test Battery suggested here improve the prediction? If expectancy tables were introduced into the system at the trial level, with or without test data, would the accuracy of the expectancy tables as a predictor of parole success be improved?

Correlations between variables can be improved by statistical corrections for restriction of range, but we need to know the range of the general population from which the sample has been drawn. If testing procedures could be introduced at some earlier stage of the system, as recommended here, would the data permit better prediction of parole success?

APPENDIX

PRIMARY REFERENCES FOR TESTS

MMPI - Minnesota Multiphasic Personality Inventory, 1042-67
Starke Hathaway and J. Charnley McKinley

Psychological Corporation
757 Third Avenue
304 E. 405th Street
New York, N. Y. 10017
\$1.75/Manual

EDS - Environmental Deprivation Scale, May, 1974
M. D. DeVine, W. O. Jenkins, A. D. Witherspoon,
E. K. DeValera, J. B. Muller, and J. M. McKee

Rehabilitation Research Foundation
P. O. Box 3587
Montgomery, Alabama 36109

Cost - \$3.00

LESS - The Law Encounter Severity Scale, May, 1974
J. B. Muller, M. D. DeVine, E. K. DeValera,
A. D. Witherspoon and J. M. McKee

Rehabilitation Research Foundation
P. O. Box 3587
Montgomery, Alabama 36109 - Cost \$3.00

WAR- Weekly Activity Record - May, 1974
W. O. Jenkins, J. B. Muller, M. D. DeVine,
E. K. DeValera, A. D. Witherspoon and J. M. McKee

Rehabilitation Research Foundation
P. O. Box 3587
Montgomery, Alabama 36109

Cost - \$3.00

MBR - The Maladaptive Behavior Record, May, 1974
Marlin Barton and W. O. Jenkins

Rehabilitation Research Foundation
P. O. Box 3587
Montgomery, Alabama 36109

Cost - \$3.00

CPI - California Psychological Inventory, 1956-69
Harrison G. Gough

Consulting Psychologists Psychological Inventory, 1956-69
Harrison G. Gough

Consulting Psychologists Press, Inc.
577 College Avenue
Palo Alto, California 94306

Cost - \$4.00/Manual
\$1.00/spec. set

Jesness Inventory, ci. 1962-66
Carl F. Jesness

Consulting Psychologists Press
577 College Avenue
Palo Alto, California 94306

Cost - \$2.00/spec set

Bender-Gestalt, 1951
Gerald R. Paschal and Barbara J. Suttel

Grune and Stratton, Inc.
757 Third Avenue
New York, New York 10017

Cost - \$10/manual

DAP - Draw-a-Person, 1963
William H. Urban

Western Psychological Services
12031 Wilshire Blvd.
Los Angeles, California 90025

Cost - \$6.00/manual

CAT - California Achievement Test, 1970
Ernest W. Tieg and Willis W. Clark

CTB/McGraw-Hill
330 W. 42nd Street
New York, New York 10036

Cost - \$.50/manual

TAT - Thematic Apperception Test - 1935-43
Henry A. Murray

Harvard University Press
79 Garden Street
Cambridge, Mass. 02138

Cost - \$.50/manual

Rorschach - Behn-Rorschach, Bero Test or BRT, 1941-56
Hans Zulliger and Hans Huber

Grune and Stratton, Inc.
757 Third Avenue
New York, New York 10017

Cost - \$8.00/manual

WISC - Wechsler Intelligence Scale for Children
Revised, 1974

Cost - \$5.00/manual

WAIS - Wechsler Adult Intelligence Scale, 1939-55
David Wechsler

Psychological Corporation
757 Third Avenue
New York, N. Y. 10017

Cost - \$4.00/manual

WISC - Wechsler Intelligence Scale for Children, 1949
David Wechsler

Psychological Corporation
757 Third Avenue
New York, N. Y. 10017

Cost - \$4.00/manual

WRAT - Wide Range Achievement Test, 1940-65
J. F. Jastak, S. R. Jastak, and S. W. Bijou

Guidance Associates
6516 Shirley Avenue
Austin, Texas 78752

Cost - \$2.75/spec set

GATB - General Aptitude Test Battery, 1946-70
Developed for use in U.S. Training & Employment Svs.

U. S. Government Printing Office
Washington, D. C. 20402

Cost - \$7.00(approx.) manual

PPVT - Peabody Picture Vocabulary Test, 1959-70
Lloyd M. Dunn

American Guidance Services Inc.
Publishers' Building
Circle Pines, Minn. 55014

Cost - \$10/testing materials

HTP - House-Tree-Person Technique: Revised Manual, 1966
John N. Buck

Western Psychological Services
12031 Wilshire Blvd.
Los Angeles, California 90025

Cost - \$12.50/manual

HSPQ - High School Personality Questionnaire, 1953-69
Raymond B. Cattell, Richard W. Coan and Halla Beloff

Bobbs-Merrill Co., Inc.
4300 W. 62nd Street
Indianapolis, Ind. 46268

Cost - \$1.50/spec set
\$1.25/manual

TSCS - Tennessee Self Concept, 1964-65
William H. Fitts

Counselor Recordings and Tests
Box 6184 Acklen Station
Nashville, Tenn. 37212

Cost - \$.90/spec set

Gough Socialization Checklist, 1952-65
Harrison G. Gough, Alfred B. Heibrun, Jr.

Consulting Psychologists
577 College
Los Angeles, California

Cost - \$2.50/manual
\$2.50/spec set

END