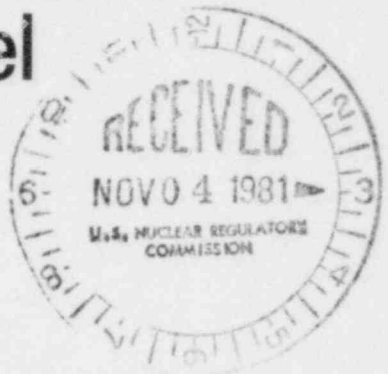
NUREG/CR-2075

Standards for Psychological Assessment of Nuclear Facility Personnel



Prepared by F. D. Frank, B. S. Lindley, R. A. Cohen

Assessment Designs, Incorporated

Prepared for
U.S. Nuclear Regulatory Commission

Available from

GPO Sales Program
Division of Technical Information and Document Control
U. S. Nuclear Regulatory Commission
Washington, D. C. 20555

Printed copy price: $6.50

and

National Technical Information Service
Springfield, Virginia 22161

NUREG/CR-2075
1S

Standards for Psychological Assessment of Nuclear Facility Personnel

Manuscript Completed: December 1980
Date Published: July 1981

Prepared by
F. D. Frank, B. S. Lindley, R. A. Cohen

Assessment Designs, Incorporated
ADI Court - 601 North Ferncreek
Orlando, FL 32803

Prepared for
Division of Facility Operations
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555
NRC FIN B1083

ABSTRACT

The subject of this study was the development of standards for the assessment of emotional instability in applicants for nuclear facility positions. The investigation covered all positions associated with a nuclear facility. Key findings were that emotional instability is a multi-dimensional concept; no single instrument by itself is capable of measuring emotional instability; few studies have been conducted in a nuclear setting aimed at determining the predictive validity of various selection instruments with respect to emotional stability; and standard criteria for evaluating instruments require careful consideration of psychometric principles and legal considerations. Conclusions reached in this investigation focused on the ingredients of an integrated selection system including the use of personality tests, situational simulations, and the clinical interview; the need for professional standards to ensure quality control; the need for a uniform selection system as organizations vary considerably in terms of instruments presently used; and the need for an on-the-job behavioral observation program. In terms of key recommendations, the selection system would vary as a function of the demands of the position, and the degree and frequency of access to vital or protected areas in the facility associated with the position. More specifically, for positions of considerable on-the-job stress, the selection system would include the Minnesota Multiphasic Personality Inventory, the Sixteen Personality Factor Questionnaire, the clinical interview, and, in the case of some positions, situational simulations. For other positions, because of a lack of on-the-job related stress, and limited access to vital or protected areas, no screening for emotional instability would be necessary. When situational simulations are to be included for a given position, these instruments would need to be specifically tailored to the given position. Research needs to be conducted on the predictive validity of the aforementioned instruments, as well as others available, within a nuclear facility setting.

CONTENTS

Contents (continued)

Contents (continued)

Contents (continued)

Contents (continued)

Contents (continued)

Contents (continued)

ACKNOWLEDGMENTS

The authors wish to express their appreciation to the numerous individuals and organizations who provided reference materials, technical guidance, and valuable suggestions relevant to our objectives.

Our sincere appreciation is extended to the various individuals within the Nuclear Regulatory Commission who assisted us in our task. Special gratitude goes to John Montgomery, Office of Policy Evaluation (formerly with the Office of Standards Development), who served as our Contracting Officer's Authorized Representative. His guidance and assistance throughout the project were vital to its successful completion. We are indebted to our Contracting Officer, Ray Gustave of the Research Contracts Branch. We appreciate the comments and suggestions provided by W.B. Brown, Chief of Safeguards Standards Branch, and Sarah A. Mullen, of the Physical Security Development Branch, in their critiques of earlier work submitted under this contract.

We are particularly grateful to the utility organizations which allowed us to visit their facilities, and to their employees, whose time and input were essential to the completion of the Job Analysis of Nuclear Facility Security Personnel,* as well as other critical tasks.

Special thanks are extended to the members of an interdisciplinary group of experts who participated in a panel meeting held for the purpose of development of standard measurement instrument criteria.** The comments, suggestions, and sharing of expertise of these individuals were extremely valuable in helping us to compile the final elements of the document.

We are indebted to other members of the staff of Assessment Designs, Inc., for their assistance in this project. Special thanks go to Cabot Jaffee, President, for his professional assistance as well as for his participation in our Criteria Development Panel Meeting. We appreciate the management support provided by Joyce Generali, Administrative Assistant. We are grateful to Robert Adams for his

---

* Appendix A comprises the Job Analysis for this project.

** Appendix B contains a list of the Members of the Standards Development Panel and their affiliations.

assistance in the compilation of technical material. Finally, we are especially indebted to our Manager of Clerical Production, James Wilson, and the members of his staff, Jacquelyn Deamer, Debra Emery, Emilia Phipps, Carol Foltz, Donna Bergeron, and Tina Waszczak, whose long hours and hard work were invaluable in the production of our reports.

We would like to express our deep appreciation to the countless individuals associated with the nuclear industry around the country, with whom we spoke during the course of our work on this project. Their helpful assistance is reflected throughout this report.

Fredric D. Frank
Barbara S. Lindley
Robert A. Cohen

December, 1980
Orlando, Florida

EXECUTIVE SUMMARY

Purpose: The purposes of the investigation were to define those aspects of emotionally unstable behavior, potentially demonstrable by security guards, which would pose risks to nuclear facilities; to establish and define standard criteria for the evaluation of selection techniques which purportedly predict emotionally unstable behaviors expressed on the job; using the standard criteria, to evaluate selection techniques in terms of the degree to which they satisfy the standard criteria; and to determine the degree to which these selection techniques would have applicability to other positions (beyond security positions) within nuclear facilities.

Procedure: The first step in the investigation was a careful examination of the demands associated with the jobs of security guard and security supervisor within a nuclear facility. Visits were made to nuclear sites where interviews were conducted with job incumbents, superiors, and nonsecurity management personnel. The data gathered through the on-site visits and through other source material were analyzed to produce a job analysis report detailing the demands of the security guard and supervisor positions. Particular attention during this phase was directed toward identifying behavioral on-the-job indices of emotional instability. The second major step was the conducting of a literature review which examined the merits of various instruments with respect to measuring emotional instability. Instruments reviewed included personality tests, physiological stress measures, situational simulations, weighted application blanks, selection interviews, clinical interviews, background and reference checks, life-change scales, and a unique instrument potentially applicable to our objectives. Tentative standard criteria were developed, on which basis the various instruments were evaluated. The third major step was the convening of an expert panel consisting of persons with special expertise in psychometrics, clinical psychology, industrial psychology, physiological psychology, counseling psychology, psychiatry, and law. All panel members had considerable knowledge of the nuclear industry, and perspectives ranged from full-time employment in private nuclear organizations, government nuclear installations, and consultants to nuclear organizations. The mission of the panel was to provide major inputs into the redefinition of the standard criteria, and such refinement was accomplished. The fourth major step was the re-evaluation of the instruments in terms of the extent to which each instrument satisfied the final standard criteria. The fifth major step was the selection of instruments which most satisfied the standard criteria for the prediction of emotional instability for applicants for positions within nuclear facilities. The sixth major step was the determination of future research needs.

Findings: Emotional instability is a multi-dimensional concept, consisting of a number of independent behavioral indices. No single instrument, by itself, is capable of measuring this multi-dimensional concept. There has been a paucity of research studies which have directly examined the predictive validity of instruments for measuring emotional instability as expressed behaviorally in terms of nuclear facility positions. Certain personality measures, because of existing research pertaining to their ability to measure the construct of emotional stability, are regarded as key ingredients of a selection system. Situational simulations, because of their ability to measure actual on-the-job behaviors, add additional data to a well-integrated selection system. The clinical interview provides the final linkage in the system by integrating various sources of data pertaining to the applicant. In terms of variations in the selection system, the combinations of recommended instruments to be used for different positions would vary somewhat on the basis of the following factors: 1) differences in the nature of job demands, and 2) different degrees of access to vital or protected areas. Standards developed which are applicable to individual instruments include major consideration of: 1) The psychometric elements of reliability and validity; 2) Compliance with legal issues, labor relations, and Uniform Guidelines on Employee Selection Procedures; 3) Personal effects on applicants; 4) Reevaluation considerations; and 5) Susceptibility to faking. Standards developed which are independent of individual instruments and should apply to all instruments, as well as to the total selection procedure, include: 1) Training and qualifications of administrators; 2) Confidentiality considerations; 3) Subsequent evaluation of applicants; and 4) Other concerns in regard to applicants.

Conclusions: A selection system using several different selection techniques is necessary. This selection system needs to incorporate traditional personality testing, plus, in the case of certain positions, the use of situational simulations, supported by the carefully conducted clinical interview. To ensure quality control, it is essential that professional standards be developed, which would serve to guard against the mis-application of such measurement strategies. Presently, based on the review conducted as part of this investigation, nuclear facilities vary considerably in the selection systems used, and this speaks to the need for a uniform selection system. Very few organizations have in place the integrated selection system which is advocated in this document. To produce a more complete system, attention should be directed toward developing and implementing on-the-job behavioral observation programs to supplement information obtained during the hiring process. With respect to grievances or legal ramifications pertaining to selection systems, there are two major considerations. These center

around the actual selection instruments used, and the credibility of the professionals who are involved in the process. Please refer to Chapter 7 for a discussion of instrument standards and to Chapter 10 for standards of professionals.

Recommendations: The selection system for positions associated with considerable on-the-job stress should incorporate the use of the following instruments: the Minnesota Multiphasic Personality Inventory, the Sixteen Personality Factor Questionnaire, situational simulations (where appropriate) and the clinical interview. Selection system elements for less stress-oriented positions (as defined by position demands and access to protected or vital areas) would not be as full-scale; and, for some positions, no screening for emotional instability would be necessary. Professional standards should be implemented to ensure that professionals are qualified for carrying out these measurement practices within a nuclear setting, and for providing guidance to the user organization. An appellate process should be in place to allow for redress by the unsuccessful applicant. Data pertaining to successful applicants, who evidence some signs of emotional instability, may be made available to the prospective supervisor as part of an on-going behavioral observation program. Developmental recommendatons, at the request of the unsuccessful applicant, should be made available to him/her. Finally, research needs to be conducted to further determine the predictive validity of instruments for the prediction of emotional instability in nuclear facility positions.

2. INTRODUCTION

2.1 Objective

On September 13, 1979, the U.S. Nuclear Regulatory Commission awarded a contract to Assessment Designs, Inc. (ADI) to conduct a technical study focusing on the development of standards for psychological assessment of nuclear facility security personnel. The primary objectives of this investigation were: (1) to define those aspects of emotionally unstable behavior which would pose risks to nuclear facilities; (2) to establish and define standard criteria for the evaluation of selection techniques under consideration; and 3) to identify and critique psychological tests and other evaluation techniques which would potentially detect these risk factors in security personnel job applicants. On December 14, 1979, the contract was extended to incorporate all positions within a nuclear facility.

There are several purposes for investigating and implementing optimal selection procedures. These include:

- To protect against the threat of a catastrophic nuclear accident which would endanger society as a whole.
- To devise methods to accurately measure and predict an applicant's ability to tolerate monotonous tasks, yet remain vigilant enough to satisfy job responsibilities both in terms of conducting routine tasks as well as effectively responding to crisis situations.
- To ensure that applicants are able to cope with the stresses of the job in both normal and emergency situations so that they will not be subject to personal effects, such as physical and/or emotional problems stemming from overreactivity to job demands.
- To ensure the preservation of reliable power sources.
- To preserve the proper image of individual nuclear facilities.
- To reduce the effects of employee turnover on both nuclear facilities and individual employees involved.
- To protect the large capital investments in nuclear facility equipment made by specific organizations.

2.2 Scope and Limitations

This study represented both an attempt to develop standard criteria for evaluation of selection instruments, as well as

an examination of the potential applicability of specific instruments on the basis of adherence to these criteria. Our objective was initially limited to procedures relating to screening for the presence of emotional instability indices in selection procedures for the specific positions of nuclear security guard and security supervisor, and was later extended to all positions.

It must be recognized that no system which attempts to detect and predict a human trait as complex as emotional instability will be totally accurate. Thus, the selection system recommendations included in this report cannot be expected to screen out all unstable or potentially unstable applicants. Our efforts have been directed toward developing a system as efficient and accurate as possible, which would significantly improve upon testing systems, as well as provide for more uniformity in nuclear facility selection procedures throughout the country. This project reflects a keen appreciation of the need to screen out emotionally unstable persons balanced against concerns of inappropriately labeling an individual as emotionally unstable.

2.2.1 Position Delineations for Security Personnel

At various nuclear facilities the formal position title of security guard is used to refer to different, but similar, jobs. Therefore, the following job titles will be used when referring to security employees. This distinction is based upon whether or not the security employee supervises other employees.

JOB TITLE	JOB DESCRIPTION
Security Guard	A security employee who does not supervise other security employees.
Security Supervisor	A security employee who does supervise other security employees.
Security Personnel	All security employees, both guards and supervisors.

One point of clarification should be made:

- This document does not differentiate between armed and unarmed security employees (security guards and security watchmen, respectively). It is the opinion of the authors that applicants for both types of positions be screened for emotional instability.

2.2.2 Supplemental Information Relative to the Objective

Since little research has addressed the specific occupation of nuclear security personnel, we investigated data relating to selection procedures for similar occupational groups as a preliminary aid in comprising our report. This information is contained in Appendix C (Relevant Research).

2.2.3 Data Limitations

The information contained in the Job Analysis completed for this contract (see Appendix A) was primarily obtained through field work, while the majority of the individual instrument reviews was based upon a literature review. A great deal of telephone contact with resource persons was conducted to clarify and update these data as much as possible. It was our conclusion that there was a paucity of criterion-related validity studies on the instruments reviewed, and thus it was difficult at times to distinguish among the instruments based on available criterion-related validity studies.* This limitation is reflected in Chapter 6. (Reviews of Measurement Techniques/Instruments).

2.3 Informational Sources

The information used in the compilation of this document was gathered from a variety of sources, including:

- Government Documents
- Dissertation Abstracts International
- Unpublished manuscripts
- Medical books and journals
- Psychological books and journals
- Journal Supplement Abstract Service of the American Psychological Association
- Psychological test publishers
- Legal publications

---

* See Chapter 7 for information regarding the criticality of criterion-related validity studies for the validation of selection instruments. Chapter 10 also addresses this issue in terms of research needed.

- The Equal Employment Opportunity Commission (EEOC) publications

- Personal communications with interdisciplinary experts with experience related to the nuclear industry.

- Personal visits to, and communications with, nuclear facilities located in the United States and Canada

- Personal visit to the National Security Agency

- Personal visit to the Defense Nuclear Agency

- Personal visit to the National Aeronautics and Space Administration (NASA)

- Communications with private security guard agencies

- Personal visits to professional individuals/organizations (such as private consulting psychologists and psychiatrists) who currently conduct screening procedures for nuclear facilities.

- Communications with other Federal Agencies such as the Federal Aviation Administration.

- Communications with law enforcement organizations.

3. METHOD

3.1 Overview of Approach

There were several phases and a number of activities included in the development of this document. The following list of basic procedural steps was initiated subsequent to contract award:

(a) Introductory meetings with NRC officials to establish formal objectives.

(b) On-site visits to nuclear facilities to gather information pertaining to the Job Analysis of Nuclear Security Personnel (see Appendix A).

(c) Establishment of working definition of "emotional instability" in terms of behaviors/behavioral patterns that would pose risks to nuclear facilities.

(d) Establishment of measurement and other additional relevant considerations to be used in evaluating the relative merit of individual instruments.

(e) Review of selection procedure data available on other occupational groups with similar risk factors.

(f) Review of Governmental agency regulations applicable to security personnel selection.

(g) Review of potential individual selection instruments by generic categories.

(h) Panel meeting of interdisciplinary experts (see Appendix B) held at Assessment Designs, Inc. offices, for purpose of gathering input to be used in development of standard criteria for instrument evaluation.

(i) Formal establishment of standard criteria for instrument evaluation and consideration.

(j) Recommendation of combination selection procedures for nuclear security personnel, based on instrument reviews and other pertinent issues (see Chapter 8).

(k) Determination of applicability of selection procedures to other positions within a nuclear facility (see Chapter 9).

(l) Recommendatons for further research.

(m) Discussion of additional considerations pertinent to task objectives.

3.2 Dimensions of Emotional Instability

The primary justification for the inclusion of psychological assessment techniques as part of a nuclear personnel selection procedure is to screen out those applicants who appear to be emotionally unstable. There are a number of difficulties in determining the appropriate psychological assessment techniques. These are:

- Emotional instability is a hypothetical construct which implies that it is abstract and not directly observable.
- There are numerous elements and behavioral tendencies which might lead a professional to evaluate an individual as being emotionally unstable.
- Emotional instability may be manifested by an individual in certain contexts or situations and not in others; thus, it is difficult to make a general statement regarding a person's level of emotional stability.

Due to the points just mentioned, it was necessary as an initial step in our project, to devise a working definition of emotional instability that would minimize some of these difficulties. Our approach involved the following elements: (1) The term "emotional instability" was broken down into a number of dimensions reflective of specific kinds of emotional difficulties; (2) Using job analysis information pertaining to the security positions, we attempted to further define the dimensions in terms of behaviors and behavioral patterns that would pose risks to the safety and security of nuclear facilities; (3) We classified our dimensions into two main categories: a) those traits and behavioral tendencies indicative of individuals who behave inappropriately regardless of the specific situation*, and b) those traits and behavioral tendencies exhibited by persons who normally behave appropriately but tend to react inappropriately to stressful or emergency situations.

* Although our dimensions are defined in behavioral terms, these dimensions would be generally comparable to clinical syndromes such as personality disorders and various types of psychopathology.

The following sub-section lists the dimensions of emotional instability derived from the previously discussed approach.

3.2.1 Behaviors and Behavioral Patterns Associated with Risks to Nuclear Facility Security

3.2.1.1 Behavioral Patterns Indicative of Inability to Appropriately Respond to Stressful/Crisis Situations

(1) Immediate or Short-Term Reactions to Crisis Situations

- Responds impulsively with inappropriate actions
- Freezes or becomes incapacitated
- Retreats from the situation
- Does not promptly or effectively communicate incident to other personnel who should take remedial action
- Displays a startled reaction or begins crying
- Places top priority on defending his/her innocence regarding the situation.
- Shows signs of physiological reactivity such as trembling, sweating, dizziness, heart palpitations, shortness of breath, or fainting spells

(2) Reactions to Long-Term Effects of Accumulated Stress

- Exhibits deteriorating performance
- Develops mood changes
- Exhibits constant worrying
- Becomes hypersensitive to comments of others.
- Complains of subjective feelings of tension
- Complains about pressures at work as well as home, family, financial status, etc.
- Exhibits decreased frustration tolerance

- Shows signs of developing substance dependency or abuse
- Appears chronically fatigued
- Calls in sick frequently
- Develops psychosomatic symptoms such as hypertension, gastric ulcers, migraine headaches, etc.

3.2.1.2 Behavioral Patterns Indicative of Generalized Emotional Instability

(1) Hostility Toward Authority

- Refuses to follow orders
- Exhibits screaming, obscenities, violence, arguments, or temper tantrums when questioned by superiors
- Shows arrogant and critical attitude toward company
- Violates standard operating procedures
- Does not follow appropriate chain of command
- Refuses to accept help from others
- Refuses to adhere to safety precautions
- Becomes easily agitated
- Provides false or inaccurate information when questioned

(2) Illegal and Antisocial Behaviors

- Steals from organization
- Vandalizes facilities
- Engages in sabotage
- Intentionally provides inaccurate information to co-workers and superiors

(3) Irresponsibility

- Is careless in performing duties
- Is frequently tardy or absent

- Appears unconcerned with disciplinary measures
- Does not complete assignments
- Plays pranks on others on the job
- Chooses easiest or most apparent alternative
- Conducts personal business while on duty
- Acts impulsively

(4) Dependent Behavioral Patterns

- Is overly fearful of radiation exposure when no real danger is present
- Is unable to make own decisions and needs explicit instructions
- Shows signs of extreme timidity on the job
- Denies mistakes whenever possible, and makes excuses for proven mistakes
- Exhibits excessive need for approval

(5) Interpersonal Skill Deficiencies

- Shows lack of proper assertion
- Tends toward social isolation or withdrawal
- Is unable to effectively engage in casual or formal conversation
- Is unable to effectively transmit necessary information
- Stutters when trying to speak to others

(6) Deficiencies in Vigilance

- Displays low boredom tolerance
- Sleeps on the job
- Tries to create excitement
- Is inattentive to job duties

(7) Emotional and Thought Disturbances

- Shows no emotion at all
- Is overemotional (laughs, cries, becomes upset over minor things, etc.)
- Has insomnia
- Develops changes in appetite
- Appears disoriented in time and space
- Is quite forgetful and has memory lapses
- Displays recurrent mood swings, from severe depression to extreme euphoria
- Displays lack of attention to personal appearance
- Exhibits excessive suspiciousness
- Expresses sensory hallucinations
- Displays difficulty in comprehending and responding to questions
- Creates and uses meaningless words or phrases
- Displays emotional responses which are inappropriate to the situation
- Exhibits delayed reaction time
- Displays decline in intellectual functioning

4. GOVERNMENTAL AGENCY REGULATIONS APPLICABLE TO NUCLEAR FACILITY PERSONNEL SELECTION

4.1 Uniform Guidelines on Employee Selection Procedures (1978)

4.1.1 General Objectives of Guidelines

On August 25, 1978, the Equal Employment Opportunity Commission (EEOC), Civil Service Commission (CSC), Department of Labor (DOL), and Department of Justice (DOJ), jointly issued the Uniform Guidelines on Employee Selection Procedures, referred to as the Guidelines, to be effective as of September 25, 1978. The purpose of these Guidelines was to establish a unified Federal position in the area of prohibiting discrimination in employment practices on grounds of race, color, religion, sex, or national origin. Prior to the Guidelines, the EEOC, CSC, DOL, and DOJ were all responsible for administering and monitoring equal employment regulations, and two different sets of guidelines existed (Ref. 1).

In an attempt to end the confusion that existed with regard to Federal policies on employment practices, the four regulatory agencies adopted a uniform set of guidelines by which they would all abide.

In order to clarify the content of the Guidelines, the EEOC, CSC, DOL, and DOJ published a set of questions that were commonly asked with regard to the Guidelines, and answers to these questions (Ref. 2). One of the most frequently asked questions is, "Who is covered by the Guidelines?" The answer given to this question is that:

> The Guidelines apply to the Federal Government with regard to Federal employment. They apply to most private employers who have 15 or more employees for 20 weeks or more in a calendar year, and to most employment agencies, labor organizations and apprenticeship committees. They apply to state and local governments which employ 15 or more employees, or which receive revenue sharing funds, or which receive funds from the Law Enforcement Assistance Administration to impose and strengthen law enforcement and criminal justice, or which receive grants or other Federal assistance under a program which requires maintenance of personnel standards on a merit basis.

> They apply through Executive Order 11246 to contractors and subcontractors of the Federal Government and to contractors and subcontractors under federally assisted construction contracts. (Ref. 2).

Another frequently asked question is "Do the Guidelines apply only to written tests?" The answer is:

> No. They apply to all selection procedures used to make employment decisions, including interviews, review of experience or education from application forms, work samples, physical requirements, and evaluations of performance. (Ref. 2).

4.1.2 Definitions of Key Terms Contained in Guidelines

For the sake of clarity, some of the terms used in the Guidelines have been specifically defined. These terms and their meanings with regard to the Guidelines are:

(1) Adverse impact. A substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group.

(2) Substantially different rate of selection. The agencies have adopted a rule of thumb under which they will generally consider a selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5ths) or eighty percent (80%) of the selection rate for the group with the highest selection rate as a substantially different rate of selection. This "4/5ths" or "80%" rule of thumb is not intended as a legal definition, but is a practical means of keeping the attention of the enforcement agencies on serious discrepancies in rates of hiring, promotion and other selection decisions.

(3) Compliance with these guidelines. Use of a selection procedure is in compliance with these guidelines if such use has been validated in accord with these guidelines (as defined below), or if such use does not result in adverse impact on any race, sex, or ethnic group, or, in unusual circumstances, if use of the procedure is otherwise justified in accord with Federal law.

(4) Content Validity. Demonstrated by data showing that the content of a selection procedure is representative of important aspects of performance on the job (see Appendix D).

(5) Construct Validity. Demonstrated by data showing that the selection procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important for successful job performance (See Appendix D).

(6) Criterion-Related Validity. Demonstrated by empirical data showing that the selection procedure is predictive of or significantly correlated with important elements of work behavior (See Appendix D).

4.1.3 Implications for Selection Procedures Resulting from Evidence of Adverse Impact

One of the most significant changes in employment practices that was brought about by the Guidelines is that selection procedures do not have to be validated. That is, if adverse impact does not exist, the selection procedure does not have to be validated. According to the published questions and answers, the bottom line on adverse impact, affirmative action, and validation is:

> Although validation of selection procedures is desirable in personnel management, the Uniform Guidelines require users to produce evidence of validity only when the selection procedure adversely affects the opportunities of a race, sex, or ethnic group for hire, transfer, promotion, retention or other employment decision. If there is no adverse impact, there is no validation requirement under the Guidelines. (Ref. 2).

Thus, unless adverse impact exists, employers do not have to validate their selection procedures. However, unless an employer is positive that adverse impact does not exist and will not exist in the future, it may be advisable to validate the selection procedure. The three types of validity strategies recognized by the Guidelines are content validity, criterion-related validity, and construct validity.

4.2 Nuclear Regulatory Commission Guidelines

4.2.1 ANSI N18.17-1973 (ANS-3.3) Industrial Security for Nuclear Power Plants

ANSI N18.17-1973 (Ref. 3), approved on July 18, 1973, stated minimum provisions for making determinations regarding the

acceptability of candidates for nuclear power facility employment and the continuing acceptability of employees with regard to their trustworthiness, emotional stability, and behavioral competency. These provisions are:

(1) an investigation, either prior to employment or prior to assignment to a position allowing access without escort, to disclose adverse character traits that might bear on his/her abilities or motivation to discharge his/her duties in a responsible manner

(2) examination by a licensed psychiatrist or physician, or other person professionally trained to identify aberrant behavior, either prior to employment or prior to assignment to a position allowing access without escort, for the purpose of observing and disqualifying persons displaying indications of emotional instability such that there is reasonable doubt the person could discharge his/her duties in a competent manner.

(3) continued observation of all employees and appropriate corrective measures by responsible supervisors for indications of aberrant behavior of personnel in the course of performance of their duties.

4.2.2 ANSI N546-1976 (ANS-3.4): Medical Certification and Monitoring of Personnel Requiring Operator Licenses for Nuclear Power Plants

ANSI N546-1976 (Ref.4) outlines the health requirements and disqualifying conditions applicable to nuclear facility personnel requiring operator licenses. This standard applies to requirements for both initial selection and continued monitoring of licensed operators. The provisions of this standard regarding mental qualifications state that an established history or clinical diagnosis of any of the following conditions constitutes disqualification of the applicant or employee:

(1) Any psychological condition which could result in impaired alertness, judgment, or motor ability.

(2) A personality disorder severe enough to have been displayed by overt actions.

(3) A past suicide attempt.

(4) A history of psychosis

(5) Alcoholism

(6) Drug Dependence

(7) Presence or history of any other clinically significant psychological disorder in which the condition or its treatment could hamper safe performance of all operator duties

4.2.3 Title 10, Code of Federal Regulations, Part 73

The Nuclear Regulatory Commission establishes official regulations which apply to many aspects of the design, construction and operation of nuclear facilities. These statutes are included in the U.S. Code of Federal Regulations, and take precedence over American National Standards published by the American Nuclear Society.

10CFR Part 73, Appendix B (Ref. 5), outlines general criteria for security personnel. Section I.B (2) addresses the mental qualifications necessary for security personnel. These provisions include the following:

(1) Security personnel shall demonstrate mental alertness, be capable of exercising good judgment and implementing instructions, and possess sensory and expressive capabilities sufficient to permit accurate communication by written, verbal, audible, visible, or other signals required by assigned job duties.

(2) In addition to meeting the requirements listed above, armed individuals and central alarm station operators shall undergo professional evaluations to assure that they display no evidence of emotional instability that would interfere with assigned job duties. These determinations shall be made by licensed psychologists, psychiatrists, physicians, or other persons professionally trained to identify emotional instability.

(3) The licensee shall make provisions for continued observation of security personnel as a means of detection of indices of emotional instability in employees engaged in routine job duties. Those individuals identified by their supervisors as displaying unstable tendencies shall undergo evaluation by a licensed, trained person for verification purposes.

4.2.4 Title 10, Code of Federal Regulations, Part 55

Another NRC Federal regulation applicable to nuclear facility personnel selection is 10CFR, Part 55.11 (Ref. 6). Paragraph A (1) of Part 55.11 addresses requirements for approval of operator position employment applications. This regulation states that an applicant for an operator position must not show evidence of any medical disorder which might cause inadequate performance of required job duties. The specific disorders related to emotional instability which would result in disqualification include, "insanity or any other mental condition which might cause impaired judgment or motor coordination."

4.3 References - Section 4.

1. Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice, "Uniform Guidelines on Employee Selection Procedures," Federal Register, Vol. 43, No. 166, 1978, 38289-38315.

2. Equal Employment Opportunity Commission, Office of Personnel Management, Department of Justice, Department of Labor, and Department of the Treasury, "Adoption of Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures," Federal Register, Vol. 44, No. 43, 1979, 11995-12009.

3. "American National Standard for Industrial Security for Nuclear Power Plants," ANSI N18.17-1973 (ANS 3.3). Available from American Nuclear Society, 555 North Kensington Avenue, La Grange Park, IL. 60525.

4. "American National Standard for Medical Certification and Monitoring of Personnel Requiring Operator Licenses for Nuclear Power Plants," ANSI N546-1976 (ANS 3.4). Available from American Nuclear Society, 555 North Kensington Avenue, La Grange Park, IL 60525.

5. Title 10, Code of Federal Regulations, Part 73, "Physical Protection of Plants and Materials," Available from Government Printing Office, Washington, D.C.

6. Title 10, Code of Federal Regulations, Part 55, "Operators' Licenses," Available from Government Printing Office, Washington, D.C.

5. PRELIMINARY FACTORS PERTINENT TO CRITICAL REVIEW OF MEASUREMENT INSTRUMENTS

5.1 Introduction

Prior to presenting our review of measurement instruments under consideration, several points of clarification regarding evaluation factors and terminology are necessary. These factors are presented in the following sub-sections.

5.2 Measurement Techniques/Measurement Instruments: Definition and Distinction

When referring to the instruments which are used to select job applicants, a distinction must be made as to whether the reference is to the general class of instruments, or to the specific instruments themselves. For the purposes of this review, the general class of instruments will be referred to as measurement techniques, while the specific instruments will be referred to as measurement instruments.

For example, if the measurement technique is personality tests, examples of measurement instruments would be the Minnesota Multiphasic Personality Inventory (MMPI) and the California Psychological Inventory (CPI).

The validity, reliability, etc. of measurement instruments which fall into the same category of measurement techniques may vary; therefore, separate descriptions and discussions of each instrument within a measurement technique will be carried out where appropriate.

5.3 Factors and Considerations Applied to Evaluation of Measurement Instruments/Techniques

The following factors were evaluated in regard to each measurement instrument/technique reviewed in Chapter 6 of this report.

(1) Measurement Considerations*
   (a) Reliability
   (b) Validity

* Appendix D contains definitions of the measurement and other technical terms used as evaluation factors.

(2) Additional Relevant Considerations
   (a) Cost of Development and Administration
   (b) Administrator: Training and Qualifications
   (c) Administration of Measurement Technique: Time and Difficulty
   (d) Equipment/Materials Needed for Administration
   (e) Personal Effects on Applicants
   (f) Compliance with Legal Issues and EEOC Guidelines
   (g) Confidentiality of Measurement Technique Results
   (h) Susceptibility to Faking
   (i) Labor Relations Considerations
   (j) Reevaluation Considerations
   (k) Applicability of Measurement Technique to Selection of Nuclear Facility Personnel with Regard to Emotional Instability

6. REVIEWS OF MEASUREMENT TECHNIQUES/INSTRUMENTS

6.1 Introduction

The following chapter addresses the specific techniques and instruments applicable to the measurement and prediction of emotional instability. The literature was extensively reviewed to determine available measures of emotional instability, particularly with respect to the demands associated with nuclear facility personnel positions. This investigation revealed the following categories of techniques, each of which will be discussed individually: personality tests, physiological measurement techniques, situational simulations, weighted application blanks, clinical interviews, selection interviews, background and reference checks, and life-change scales. It is important to remember that these techniques yield primarily temporal measures which may not show stability over time. This factor bears implications for the need for periodic reevaluation or on-the-job observation of those already hired (see Chapter 10 of this document for a more thorough discussion of this issue).

In addition to the previously listed categories of measurement techniques, we reviewed one unique type of instrument, the Zero Input Tracking Analyzer/Auxiliary Distraction Task (ZITA/ADT). This instrument, although less widely used than the instrument types previously mentioned, may have potential use in a selection procedure within a nuclear facility setting (see Section 6.10 for a review of this instrument).

It is recognized that there are some other unique measurement devices which may have promise; however, our review did not reveal any such instruments which have undergone sufficient investigation to be seriously considered at this point.

6.2 Personality Tests: Introduction to Specific Tests

Personality tests are most widely used in clinical settings; however, they are frequently used in employment settings. The primary objective of the various types of personality tests is to provide information regarding an individual's emotional, motivational, attitudinal and interpersonal characteristics (Ref. 1). Such instruments have been in existence since the nineteenth century.

The first instrument designed to measure personality and emotional traits was the "Free Association Test" (Ref. 1). Since that time, there have been many personality tests developed. Categories of personality tests are typically based upon a particular theory of personality, which is used to explain human behavior. Some examples of these categories of tests are self-report inventories, projective techniques, personality questionnaires, and sentence completion tests.

In discussing tests that a e designed to measure concepts such as personality types, motivational traits, and relative degrees of emotional stability, it must be pointed out that these concepts are examples of psychological constructs. A construct is an abstract term or concept, which is developed as part of a theoretical framework to explain certain events occurring in nature, such as observable behaviors (Ref. 2). Thus, a construct is inferred and cannot be directly observed. Instead, it must be measured by instruments presumed, and then validated to be representative of that construct.

The development and research of personality tests have undergone considerable growth since the development of the first systematic personality tests during World War II. There are now several hundred personality tests available in a number of different formats, for use in various settings and for a variety of purposes. Their widespread use is largely due to relative ease of administration, low cost and in some cases, standardized test forms and manuals.

Although subject to various applications, personality testing remains primarily a tool for use in clinical appraisals.

In our review of specific personality tests, it will be noted that a number of available tests are not discussed. This exclusion is based on a preliminary consideration of such factors as available and supportive data pertaining to issues of reliability, standardization of testing conditions, occupational norms, indices of emotional instability, and validity within the nuclear facility setting. Additionally, certain collective groups of personality tests, such as projective tests, have been omitted due to a lack of hard supportive data and to significant variations employed in test administration, scoring and interpretation.

Some of the specific tests to be reviewed in the following sub-sections of Section 6.2 can be computer scored while some can be both scored and interpreted by machine. In considering the convenience of these computerized services, it is important to exercise caution in regard to their utility. Despite a tendency to view information presented on a computerized print-out as statements of fact, it must be remembered that in regard to personality tests, such information is not to be considered any more definitive than a hand-scored and interpreted personality test profile. However, some computerized services are capable of offering more job-related interpretive information, and may control inconsistencies in human interpretation. These factors should be weighed in the consideration of using a computerized interpretation of a particular personality test.

6.2.1 Minnesota Multiphasic Personality Inventory (MMPI)

6.2.1.1 Overview

The MMPI, initially published by Hathaway and McKinley in 1943, is a True-False Inventory consisting of 550 items. This instrument was developed to furnish an objective evaluation of some of the major personality traits which influence individual and interpersonal adjustment. The first form of the test yielded scores on nine scales developed for clinical use and was based on abnormal personality traits. The original scales were found to also be applicable to the normal range of behavior, so they were retained with their original labels, which were designated in terms of pathological conditions. The revised form of the test (Ref. 3), which is currently in use, includes the following ten clinical scales and three validity scales:

Clinical Scales	Validity Scales
(1) Hs (hypochondriasis)	(1) L (lie)
(2) D (depression)	(2) F (validity)
(3) Hy (hysteria)	(3) K (correction)
(4) Pd (psychopathic deviate)	
(5) Mf (masculinity-femininity)	
(6) Pa (paranoia)	
(7) Pt (psychasthenia)	
(8) Sc (schizophrenia)	
(9) Ma (mania)	
(10) Si (social introversion)	

The validity scales are included in the test to check for carelessness, misunderstanding of test items, and response-sets in which the individual tries to deliberately make him/herself look good or bad in completing the test. Interpretation of the MMPI profile implies looking at the pattern produced by all thirteen scales, rather than viewing one scale in isolation. There are two major coding systems employed in the interpretation of MMPI profile patterns (Refs. 4 and 5). The variations between these two coding systems may present inconsistencies in profile interpretation and subsequent behavioral prediction. Therefore, users of this instrument should clearly specify the coding system employed in their interpretations. In addition, the choice of systems should remain consistent for a given application of the instrument.

The MMPI is the most widely used personality inventory (Ref. 1). Although its primary use is in clinical settings, a number of organizations use this instrument in their personnel selection procedures. Several computer scoring and interpretation services now offer mechanized systems for rapid processing of this test, making it an attractive instrument to organizations which screen large numbers of applicants. Users of such services should weigh the relative merits and

precautions regarding the use of computerized test interpretations as previously discussed in Section 6.2.

6.2.1.2 Relevant Considerations

6.2.1.2.1 Reliability

With respect to the MMPI, traditional evaluation studies have centered on test-retest reliability, while some research has also addressed internal consistency reliability.

Test-retest reliability studies by Hathaway and McKinley, Cottle and Holzberg, and Alessi (cited in Ref. 3) revealed test-retest reliability coefficients for the various scales ranging from .46 to .91 with a mean of .74 for normal subjects and from .59 to .93 with a mean of .76 for psychiatric patients. Most of the lower reliabilities have been found on scales which measure fluctuating personality variables such as depression, which often varies as a function of real-life situational problems.

Internal consistency reliability studies by Dahlstrom et. al. (cited in Ref. 1) revealed particularly low reliability coefficients using the split-half method. As noted by Anastasi, however, this might be expected given the heterogeneity (i.e., differences in types) of item content.

6.2.1.2.2 Validity

With respect to the MMPI, traditional evaluation studies have centered on the criterion-related and construct validation approaches to the measurement of validity.

Spielberger (Ref. 6) reported on a number of concurrent validation studies which have been conducted in an effort to predict police performance using the MMPI. However, these studies have been conducted using a number of different combinations of MMPI scales as predictors, as well as variations in criterion measures. Thus, the results are quite variable and not uniformly supportive.

Hathaway and McKinley (Ref. 3) reported that efforts to demonstrate construct validity for the MMPI showed that a high score on a clinical scale was shown to predict the final clinical diagnosis (as designated by attending psychiatrists) in more than sixty percent of new psychiatric admissions. Anastasi (Ref. 1) stated that the construct validity of the MMPI has been gradually strengthened by accumulation of correlational empirical data regarding persons who demonstrate particular kinds of profiles.

6.2.1.2.3 Cost of Development and Administration

The MMPI is available in three forms: Group Form, Form R (which may be computer or hand-scored), and the Individual Form (which must be hand-scored). The current prices of the test materials are as follows:

Group Form:	Package of 25 reusable Inventory Booklets - $9.75 Package of 500 machine-scorable Answer Documents - $68.00
Form R:	Package of 10 reusable Inventory Booklets - $38.00 Package of 500 machine-scorable Answer Documents - $68.00
Individual Form:	Box of 500 Item Cards - $39.50 Package of 500 Recording Sheets - $58.00 Manual and Transparent Scoring Keys - $13.00

There are several organizations which perform the machine scoring and some that provide an automated interpretation of each individual profile as well. The prices for these services vary from firm to firm, generally averaging from approximately $3.00 per profile for scoring only, to about $10.00 per profile for scoring and interpretation.

6.2.1.2.4 Administrator: Training and Qualifications

Since the MMPI is considered a self-report inventory, administration does not require any specific training in psychology. However the administrator must be acquainted with the standard set of instructions provided to the subjects. Interpretation of the MMPI profile does require that the individual be a trained and qualified psychologist or psychiatrist, who is thoroughly familiar with this instrument.

6.2.1.2.5 Administration of MMPI: Time and Difficulty

The time of administration of the MMPI varies from person to person, but generally takes anywhere from forty-five minutes to an hour and a half. Assuming that the individual has an adequate reading ability so that the questions are understood, there are no inherent difficulties associated with test administration.

6.2.1.2.6 Equipment/Materials Needed for Administration

The materials needed for administration of the MMPI include the Inventory Booklet, Answer Document, and soft-lead pencils.

6.2.1.2.7 Personal Effects on Applicants

Some of the items on the MMPI might be considered objectionable to some individuals; however, this possible invasion of privacy appears to be the only potential type of detrimental effect on the applicants, and may be resolved by the applicant's choosing not to respond to such items.

6.2.1.2.8 Compliance With Legal Issues and EEOC Guidelines

Anastasi (Ref. 1) pointed out that personality tests can be expected to demonstrate significant cultural differences when members of different ethnic groups are administered the same tests. While this has been shown to occur with regard to the MMPI, some normative data on a variety of cultural and subcultural groups have now been accumulated for normative purposes (Dahlstrom, et.al., & Langon, cited in Ref. 1). The interpreter of the MMPI profile should exercise caution in evaluating the profile of any minority group member tested to prevent the possibility of adverse impact to such applicants.

Since the MMPI provides scores on a number of specific traits which may be seen as indices of the construct of emotional instability, the accumulated evidence for construct validity of this instrument appears adequate with regard to EEOC guidelines. The accumulation of normative data with regard to specific occupational groups is advisable, however, for justification of inclusion of the MMPI in specific occupational selection systems.

6.2.1.2.9 Confidentiality of MMPI Results

The American Psychological Association's Ethical Standards of Psychologists (Ref. 7) states that information obtained through psychological evaluation data should be transmitted only to persons clearly concerned with the case. If this information is to be transmitted to potential employers, the professional who conducts the evaluation should safeguard against misuse of test data, providing test interpretations (in readily understandable terms) rather than test scores, where appropriate.

6.2.1.2.10 Susceptibility to Faking

The validity scales included on the MMPI are designed, in part, to detect faking on the part of the test-taker (this faking would most likely take the form of an applicant's trying to make him/herself look good) In addition to detecting deliberate faking, the validity scales on the MMPI can be used to detect profiles reflecting an applicant who did not understand the test items. These scales were constructed from normative data, which were the same data used in developing the clinical scales of the present form. Therefore,

assuming that the MMPI profile is properly interpreted by a trained professional, most incidents of attempted faking or other response sets would be detected.

6.2.1.2.11 Labor Relations Considerations

This review did not reveal any major labor relations disputes or court decisions already resolved which centered on the use of the MMPI. There is, however, a case in process involving a private nuclear facility employee who was originally denied promotion from a non-nuclear operations position to a nuclear operations position on the basis of psychological evaluation results. This evaluation included a clinical appraisal and an MMPI. Following inconsistent recommendations based on two subsequent psychological evaluations by independent clinicians, the employee's union decided to take the case to arbitration. At the present time, no final decision has been made in this case.

Given some evidence regarding possible cultural effects with the MMPI, future labor disputes arising from discrimination based on this instrument cannot be ruled out. The efforts of on-going research to provide normative data on a variety of ethnic and racial groups for this instrument should, however, make its use more routinely acceptable with respect to subgroups within the population.

6.2.1.2.12 Reevaluation Considerations

In accordance with EEOC Guidelines, a job applicant screened out on the basis of test results should be given an opportunity for reevaluation at a later date. Provisions for retesting appear particularly important with regard to the MMPI and similar personality inventories which yield some scores that might be expected to show significant fluctuations from one testing occasion to another.

6.2.1.2.13 Applicability of the MMPI to Selection of Nuclear Facility Personnel with Regard to Emotional Instability

As previously mentioned, the MMPI is the most widely used personality inventory (including use in personnel selection procedures as well as clinical applications), and has been the subject of more research than any other personality inventory. It is one of the most comprehensive inventories available, in terms of amount of data generated for interpretation purposes, and is designed to measure all the major categories of aberrant behavior.

The reliability data on the MMPI are generally acceptable, although the majority of the studies have been conducted

using the test-retest reliability approach.* Research has demonstrated good evidence for construct validity of the MMPI, although there is a need for investigations of criterion-oriented validity using the MMPI with occupations within nuclear facilities. Normative data has been collected on various sub-groups within the population, making this instrument less likely to result in adverse impact than some of the other personality inventories available. The MMPI includes several validity scales designed to detect distortion of results on the part of job applicants.

One drawback regarding the use of the MMPI for employment (as opposed to clinical) purposes is its reliance on psychiatric classification labels for the test scales. Although the scales include ranges representative of the "normal" population, the psychiatric terminology can be misleading. Therefore, interpretive reports for use in personnel selection should omit such labels and instead provide descriptive information in behavioral terms regarding an applicant's profile.

Since the MMPI is so well-established and is familiar to most professionals, there would be less need for special training to incorporate this instrument into a nuclear facility personnel selection procedure.

6.2.2 California Psychological Inventory (CPI)

6.2.2.1 Overview

The CPI is a widely-used personality inventory. This test, first published by Gough in 1957, is one of several instruments that was developed in response to the MMPI. In contrast to the MMPI, however, the CPI was originally intended to be used with "normal" or non-clinical populations.

---

* Refer to Section 7.2.1 for a discussion of reasons why further research on the MMPI (as well as personality tests in general) should be conducted using the inter-rater reliability approach.

The CPI is a true-false inventory consisting of 480 items which yield standard scores on eighteen different traits. These traits are representative of personality characteristics which Gough felt were important for interpersonal interaction, such as Sociability, Self-Acceptance, Responsibility, Flexibility, Achievement via Conformance, Achievement via Independence, etc. Three of the eighteen scales are designed to act as validity checks on the profile, as a way of evaluating test-taking attitudes. These scales are "Sense of Well Being" (based on responses by normals requested to "fake bad"), "Good Impression" (based on responses by normals instructed to "fake good"), and "Communality" (based on the frequency of highly popular responses provided). The original normative data for the CPI were gathered from a sample of 6,000 males and 7,000 females, including wide variations in age, socioeconomic level and geographic area. There have been separate norms devised for many special groups including a variety of occupational fields. In addition, the CPI has been translated into a variety of foreign languages, and some of the scales have been shown to predict actual behavior in other countries as well as they do in the United States (Ref. 8).

6.2.2.2 Relevant Considerations

6.2.2.2.1 Reliability

With respect to the CPI, traditional evaluation studies have centered on the test-retest reliability approach to the measurement of reliability.

Gough (Ref. 9) cited two test-retest reliability studies which were conducted with high school students and male prison inmates, respectively. The correlations for the prisoner groups were comparable to those generally found in personality measurement ranging from figures of .49 to .87 with a mean of .75 for the various scales. The data collected from the high school students showed lower overall coefficients for the eighteen scales, ranging from .38 to .771 with a mean of .65. These lower reliability values may be due to differing rates of maturation among adolescents, as Gough suggests. However, in comparing the two studies, it must also be noted that the time lapse between test administrations was one year for the high school students, and only seven to twenty-one days for the inmates.

6.2.2.2.2 Validity

With respect to the CPI, traditional evaluation studies have centered on the criterion-related validity and construct validation approaches.

As Gough (Ref. 9) pointed out, only certain scales of the CPI, such as the Achievement Scales, have been subjected to criterion-related validity studies, because these scales have some clear external criteria, such as grades, against which comparisons may be made. He cited one such study which compared the achievement via conformance scale with high school students' grade point averages, resulting in validity coefficients of .41 for both male and female students. A number of other studies have attempted to establish predictive criterion-related validity for the CPI with occupational groups. Despite some supportive results, none of the scales on the CPI used as predictors nor the criterion measures in these studies, specifically related to emotional instability. Thus, it cannot be said that the CPI is a validated predictor of emotional instability at this time.

There has been a great deal of research aimed at establishing the construct validity of the various CPI scales. The two basic approaches have been comparisons of CPI scale scores with subjective trait ratings by peers or superiors, and comparisons of CPI scale scores with scales on other inventories designed to measure the same constructs. Gough (Ref. 9) cited several studies using the first approach, which compared single scales from the CPI with subjective ratings. The resulting validity coefficients from the combined data ranged from .21 to .76 with a mean of .40. Gough (Ref. 9) also cited studies comparing the various CPI scales with scales on other instruments designed to measure the same or similar traits. These studies yielded validity coefficients ranging from .32 to .60 with a mean of .45.

6.2.2.2.3 Cost of Development and Administration

The current costs of the materials necessary for administration of the CPI follow immediately:

Package of 100 Reusable Question Booklets -	$25.00
Package of 500 Handscorable Answer Sheets including Profiles -	$35.00
Manual and Handscoring Stencils -	$10.50

As with the MMPI, there are several organizations which provide computerized scoring and interpretation services, available at similar prices.

6.2.2.2.4 Administrator: Training and Qualifications

The CPI is a self-administered inventory and therefore does not require any particular training or expertise. Interpretation of the CPI profile should, however, be conducted by a trained psychologist or psychiatrist who is thoroughly familiar with this instrument and personality testing in general.

6.2.2.2.5 Administration of CPI: Time and Difficulty

Time of administration of the CPI ranges from about forty-five minutes to one hour. Assuming that the subject has an adequate reading ability, no difficulties should be expected during test administration.

6.2.2.2.6 Equipment/Materials Needed for Administration

The materials needed for administration of the CPI include the Question Booklet, the Answer Sheet, and soft-lead pencils.

6.2.2.2.7 Personal Effects on Applicants

The occurrence of objectionable items on the CPI, which might be viewed by applicants as an invasion of privacy, appears to be the only potential adverse effect on the applicants.

6.2.2.2.8 Compliance With Legal Issues and EEOC Guidelines

Since considerable effort has been spent on making the CPI a cross-culturally fair testing instrument, it does not appear to have as much potential for adverse impact to applicants due to differences in socio-economic class or ethnic background as some other personality tests might pose.

This review did not reveal any court cases to date which have centered on the use of the CPI.

6.2.2.2.9 Confidentiality of CPI Results

The CPI, being a psychological test, is subject to the same Ethical Standards for Confidentiality as outlined for the MMPI and all other personality tests.

6.2.2.2.10 Susceptibility to Faking

The validity scales included on the CPI are designed to detect test-taking attitudes indicative of an individual's attempting to slant the test results in an effort to deliberately make him/herself look good or bad. These scales are comprised of items based on normative data, so significant deviations from these norms should alert the trained interpreter of the possibility of attempted faking due to motivational factors.

6.2.2.2.11 Labor Relations Considerations

This review did not reveal any actual labor relations controversies which centered on the use of the CPI. Such a dispute would usually be expected to arise only if legal, ethical or EEOC guidelines were violated.

6.2.2.2.12 Reevaluat n Considerations

In accordance with EE(Guidelines and because some of the CPI scales reflect tr s which might be expected to display some situational fluc tions, an applicant should be given the opportunity for re valuation.

6.2.2.2.13 Applicability of the CPI to Selection of Nuclear Facility Personnel with Regard to Emotional Instability

Although research has been conducted using the CPI with various occupational groups, such studies have not yet been carried out with nuclear facility employees. Despite its widespread use, supportive research, and comprehensive nature, the CPI focuses mainly on personality traits reflective of interpersonal interaction styles. While such data are informative, it is inadequate in providing data regarding idiosyncratic aspects of an individual's personality, particularly with regard to the detection of or predisposition to many psychological disorders. In addition, the CPI takes from forty-five minutes to an hour to administer, a much lengthier time than required for some other personality tests. Thus, its inclusion in a personnel selection procedure designed to screen for emotional instability does not appear to have adequate justification at this time.

6.2.3 Sixteen Personality Factor Questionnaire (16PF)

6.2.3.1 Overview

The 16PF is a forced-choice, self-report personality inventory which is designed to provide measures of overall personality traits reflective of interpersonal relations styles in the "normal" adult personality. This test was first published by Cattell, Eber and Tatsuoka in 1970, following a thirty-year research program conducted by Cattell and his colleagues. The emphasis of this research was to isolate the basic factors or traits underlying an individual's personality and behavioral response style in various situations. Cattell felt that these "source traits" were more efficient when used in combination to predict actual behavior, than were more surface types of traits or other kinds of scales (Ref. 10, p. 330). These sixteen source traits measure personality dimensions such as "Affected by Feelings vs. Emotionally Stable," "Relaxed vs. Tense," "Trusting vs. Suspicious," "Shy vs. Venturesome," "Conservative vs. Experimenting," "Group-Oriented vs. Self-Sufficient," etc.

There are five forms of the 16PF available, with the choice of form dependent upon the purpose of, and time available for testing. Forms A and B each contain 187 items and require a reading ability equivalent to that of an average seventh grader. Forms C and D have 105 questions each, and call for

a sixth-grade reading ability. Form E is designed for use with low-literacy level individuals who have reading skills that approximate the average third grader. Forms C and D are frequently used in occupational selection work. The 16PF has been translated into twenty-four foreign languages and has also been adapted for five other English-speaking cultures.

The publisher of the 16PF provides a computerized scoring service which offers a variety of types of reports. The specific type pertinent to this review is The Personal Career Development Profile (PCD Profile). This report averages four to five pages in length, and provides, in addition to a profile of the job applicant's scores, an interpretive discussion including sections describing problem-scoring patterns, patterns for coping with stressful conditions, patterns of interpersonal interaction, and personal-career development considerations. The vocabulary and narrative style of the report are worded in a manner which is designed to assist in situations where interpretation of test results to the job applicant is essential. This report is a computerized interpretation of job-related characteristics based on the individual's personality profile. It is not, however, necessarily more accurate than a human interpretation of the same profiles would be. The computerized interpretation should be handled by individuals trained in the use of the 16PF, who can exercise the same cautions in evaluating the validity of the profiles and results as if they were being processed through an actual clinical analysis.

Normative data on the 16PF for forms A,B,C, and D were collected on a total of 15,000 American adult males and females. In addition, normative data have been collected on more than fifty occupational groups and about the same number of psychiatric syndromes. Scores on the tables from these data are presented in Sten ("Standard ten") form. These scores are distributed along a ten equal-interval score range, from 1 through 10, with a mean of 5.5 and a standard deviation of 2.

6.2.3.2 Relevant Considerations

6.2.3.2.1 Reliability

With respect to the 16PF, traditional evaluation studies have centered on the test-retest and equivalent forms approaches to the measurement of reliability.

Anastasi (Ref. 1) stated that the test-retest reliability coefficients for the 16PF sometimes fall below .80 after intervals of a week or so. She suggested that the somewhat low test-retest reliability data for this test may be due to the shortness of the scales, rather than to any inherent instrument inadequacies.

The manual for the 16PF presents equivalent-form reliability coefficients from various studies between forms A and B and between forms C and D. These coefficients are derived from a variety of studies and have a mean of approximately .58.

6.2.3.2.2 Validity

With respect to the 16PF, traditional evaluation studies have centered on criterion-related and construct validity approaches.

The manual for the 16PF specifically addresses criterion validity evidence for use in industrial settings and personnel selection. In this regard, data have been collected on numerous occupational groups, including those with job elements similar to those of nuclear facility personnel, such as police and air traffic controllers. The manual provides what Cattell calls "specification equations" for predicting an individual's criterion performance in one of these specific occupations from his/her scores on the 16PF . Although more long-term research regarding such prediction is needed, this appears to be a promising beginning in predicting occupational performance as well as on-the-job indices of emotional instability.

There has been considerable research conducted on the construct validity of the 16PF using employees in stressful occupations (such as air traffic controllers) as subjects. These data have generally revealed that the 16PF can detect significant differences in the personality traits of successful employees in such occupations and those of the general public.

6.2.3.2.3 Cost of Development and Administration

The current costs of the materials necessary for administration of the 16PF are listed below:

16 PF Handbook -	$ 9.95
Package of 25 test booklets, Forms A & B -	13.50
Package of 25 test booklets, Forms C & D	13.50
Package of Answer Sheets/Profiles, Forms A & B -	55.00
Package of Answer Sheets/Profiles, Forms C & D -	55.00
Set of Handscoring Stencils, Forms A & B -	5.25
Set of Handscoring Stencils, Forms C & D -	3.00

PCD Profiles - Cost per profile is based on quantity of profiles requested:

Quantity	Cost per PCD Profile
50-99	$10.20
100-249	9.80
250+	9.45

6.2.3.2.4 Administrator: Training and Qualifications

Because the 16PF is a self-report inventory, test administration does not require any specific qualifications; however the administrator should be trained to provide the basic test instructions, as outlined in the Administrator's Manual for the 16PF. Interpretation of the 16PF profile should be conducted by a qualified psychologist or psychiatrist who has been trained in the use of this instrument. In cases where PCD Profiles are used, trained professionals should review the profiles and accompanying interpretations to examine their validity, before such profiles are reviewed for selection consideration.

6.2.3.2.5 Administration of 16PF: Time and Difficulty

Duration of administration of the 16PF ranges from approximately 50 minutes for Forms A and B, to about 30 minutes for Forms C and D. Since the various forms of the test specify the necessary reading level of the applicant, no inherent difficulties should be expected.

6.2.3.2.6 Equipment/Materials Needed for Administration

The materials needed for administration of the 16PF include the test booklet [form(s) of choice], corresponding answer sheets, and soft lead pencils.

6.2.3.2.7 Personal Effects on Applicants

The occurrence of objectionable items on the 16PF appears to be the only potential adverse effect on applicants.

6.2.3.2.8 Compliance With Legal Issues and EEOC Guidelines

Since considerable effort has been spent on making the 16PF a culturally-fair testing instrument, it does not appear to have as much potential for adverse impact on ethnic or minority group applicants as some other personality tests might pose.

In regard to legalities involving the 16PF, it is of interest to note that the State of Pennsylvania requires the 16PF as part of its Lethal Weapons Certification procedure.

6.2.3.2.9 Confidentiality of 16PF Results

Because the 16PF is a psychological test, it is subject to the same Ethical Standards for confidentiality as previously discussed for such instruments.

6.2.3.2.10 Susceptibility To Faking

Forms A, C, and D of the 16PF all have scales built in to check for validity of the profile and detect distortion or deception. These scales have been adequately established through separate normative data and have been shown to be effective in detecting distorted response sets. Form A has three such scales: "Motivational Distortion" (to check for faking good), "Faking Bad," and "Random Responses." Forms C and D contain a single Motivational Distortion (MD) scale. The faking bad and random response scales are not included in Forms C and D, since these scales are most frequently used for occupational selection, where faking bad and random response sets would probably not be encountered.

6.2.3.2.11 Labor Relations Considerations

This review revealed no labor relations controversies in which the use of the 16PF was the major issue.

6.2.3.2.12 Reevaluation Considerations

Because some of the traits measured by the 16PF might be expected to display situational fluctuations due to real-life changes, an applicant should be provided with a reevaluation opportunity, particularly if he/she were screened out or considered questionable for hiring on the basis of this test's results. Such reevaluation would be in accordance with EEOC Guidelines. This test appears particularly well-suited for reevaluation purposes, since it includes five forms.

6.2.3.2.13 Applicability of the 16PF to Selection of Nuclear Facility Personnel with Regard to Emotional Instability

The 16PF appears to offer significant potential to the area of employment screening and selection. The PCD Profile Service appears to provide interpretive information which would reveal emotionally unstable tendencies which might interfere with job performance. The normative data available on occupations with stress factors similar to those of nuclear facility personnel appear to render this instrument particularly useful as part of a nuclear facility personnel selection procedure. To enhance this applicability, normative data should be gathered on the specific occupations within the nuclear industry.

6.2.4 Gordon Personal Profile - Inventory (GPP-I)

6.2.4.1 Overview

The GPP-I consists of two separate personality inventories which are usually used as companion instruments to provide measures of some of the major personality traits which influence behavior. The Gordon Personal Profile (GPP), first published by Leonard Gordon in 1953, provides scores on four personality characteristics that influence the average person's day-to-day functioning. These four traits include: "Ascendancy" (A), "Responsibility" (R), "Emotional Stability" (E), and "Sociability" (S). There is also a measure of "Self-Esteem" (SE) which is derived from the sum of the four trait scores.

The Gordon Personal Inventory (GPI), first published in 1956, measures four additional personality traits, including "Cautiousness" (C), "Original Thinking" (O), "Personal Relations" (P), and "Vigor" (V).

Because the GPP and the GPI are normally administered together, they are available in a combined form test booklet as well as individually. The two inventories have identical formats of the forced-choice type. Each item consists of a "tetrad" of four descriptive phrases. One of the phrases in each tetrad is descriptive of one of the personality traits measured by the instrument. Of the four phrases in each item, two are generally considered to be of similar complementary value in terms of content by typical individuals, while the other two are normally seen as having similar uncomplementary values. The task of the individual taking the test is to designate which of the phrases in each tetrad he/she considers to be most like him/herself, and which he/she views as being least like him/herself. Gordon (Ref. 11) reported that this method of forced-choice format, in which the individual has to select two responses to each item which are opposites in terms of self-applicability, rather than choosing only one alternative, appears to make the instrument less susceptible to distortion by individuals trying to make a good impression.

The combined GPP-I booklet must be hand-scored. The GPP and GPI individual booklets are available in both hand and machine-scorable forms. Both the combined GPP-I form and the individual GPP and GPI forms have been translated into more than a dozen foreign languages.

The original normative data for both the GPP and the GPI were compiled on college students. In the subsequent revisions of each instrument which followed, data were also obtained on high school, clinical, industrial, and additional college populations. In total, normative data were gathered from approximately 5,000 cases for the GPP and 1,800 for the GPI,

with sampling including representatives of diverse occupations and many geographic regions. The items in each scale were evaluated and revised in an effort to provide broad applicability. Those items retained in the final forms of the test consist of those that were found to satisfactorily discriminate among the groups to which they were administered. Rather than being converted to standard scores, an individual's raw scores are compared directly to percentile rank tables for the appropriate normative group. These tables also include the means and standard deviations for each scale.

6.2.4.2 Relevant Considerations

6.2.4.2.1 Reliability

With respect to the GPP-I, traditional evaluation studies have centered on the test-retest and internal consistency reliability approaches to measuring the reliability of this instrument.

Gordon (Ref. 11) cited two studies which investigated the test-retest reliability of the GPP-I. The first of these studies compared the scores of 127 Naval enlisted men tested at the beginning and end of a twenty-nine week training program. The correlations on the eight scales of the GPP-I ranged from .50 to .79 with a mean of .67. The other study was of longer range and compared the scale scores of members of three consecutive classes of optometry students. Each class (consisting of an average of 54 students) was tested at the beginning of the first year and retested at graduation time almost four years later. The test-retest reliability coefficients for all three classes on the eight GPP-I scales ranged from .47 to .68, with a mean of .54.

Several studies of internal consistency reliability were reported by Gordon (Ref. 11). One study, which was based on a sample of 92 college students, reported split-half reliability coefficients ranging from .86 to .89 for the GPP scales, with a mean of .88. A similar study of 168 college students demonstrated split-half reliability coefficients ranging from .80 to .83, with a mean of .82 for the GPI scales.

A study of 218 male managers in a public utility utilized the coefficient alpha, another measure of internal consistency reliability. For the GPP scales, these reliability co-efficients ranged from .82 to .85, with a mean of .83. The GPI was administered to the same sample, with the results yielding coefficient alpha reliabilities with a range of .81 to .83 and a mean of .82.

6.2.4.2.2 Validity

With respect to the GPP-I, traditional evaluation studies have centered on the construct validation approach to the measurement of validity.

The two major methods of establishing construct validity for the GPP-I have been through correlating an individual's GPP-I scale scores with ratings of the same trait by peers and others, and by correlating the scores with other personality inventories designed to measure the same traits.

Braun, Alexander, and Weiss (Ref. 12) conducted a study using the GPI with a sample of female college students who lived together in two small groups of thirteen and seventeen each. The total of thirty subjects were administered the GPI in group settings. One week later, each student was asked to rate the other members of her group on each of the four traits included on the test. The validity coefficients resulting from correlations of the test scores with peer ratings ranged from .39 to .58, with a mean of .50. In another study conducted by Bravo-Valdivieso (cited in Ref. 11), fifty-seven seminarians were administered a Spanish translation of the GPP. Each of the subjects was subsequently rated on each of the four Profile traits by three superiors. These ratings were conducted independently and then pooled for each subject. Correlations between the scale scores and the ratings resulted in validity coefficients ranging from .21 to .58, with a mean of .40.

The GPP-I scale scores have been compared with certain scales on a variety of other personality inventories. Gordon (Ref. 11) reported a study conducted on 123 Civil Service employees who were administered the GPP-I and the Eysenck Personality Inventory (EPI) (Ref. 13). Correlations were computed between GPP-I scores and the Extroversion and Neuroticism scales, which are the major scales on the EPI. As would be expected in terms of personality theory, this comparison yielded some negative correlations, such as between Emotional Stability (GPP-I) and Neuroticism (EPI), i.e., the more emotionally stable an individual, the less neurotic he/she is. The absolute values resulting from all combinations of correlated scores ranged from .05 to .58, with a mean of .30. In another study reported by Gordon, 1978, the GPP-I scale scores of 160 Naval enlisted men were compared with all of the scales on the Guilford-Zimmerman Temperament Survey (GZTS) (Ref. 14). This test is presumed to measure several traits similar to those included on the GPP-I. The absolute values of all combinations of correlated scores ranged from .00 to .65, with a mean of .23.

6.2.4.2.3 Cost of Development and Administration

The current prices of the materials necessary for administration of the GPP-I are as follows:

Package of 35 GPP-I Booklets -	$19.75
Package of 35 GPP Booklets -	10.00
Package of 35 GPI Booklets -	10.00
Package of 35 Answer Documents for Hand or Machine Scoring - GPP -	8.00
Package of 35 Answer Documents for Hand or Machine Scoring - GPI -	8.00
Handscoring Keys - GPP	5.25
Handscoring Keys - GPI -	5.25
Computerized Scoring (per test) -	1.15

6.2.4.2.4 Administrator: Training and Qualifications

No particular qualifications are necessary for administration of the GPP-I, although the individual should be trained to give the instructions for the test in a thorough and competent manner according to the guidelines in the manual. Interpretation of the GPP-I scores should be conducted by a qualified psychologist or psychiatrist who has been thoroughly trained in the use of the test.

6.2.4.2.5 Administration of the GPP-I: Time and Difficulty

Either the GPP or the GPI individual forms can be completed in approximately seven to fifteen minutes. When the combined form is administered, twenty to twenty-five minutes are usually adequate for completion. A minimum equivalent of a seventh-grade reading level is necessary for comprehension of the test items.

6.2.4.2.6 Equipment/Materials Needed for Administration

The materials needed for administration of the GPP-I include the appropriate test booklets, answer documents, and soft-lead pencils.

6.2.4.2.7 Personal Effects on Applicants

The occurrence of objectionable items on the GPP-I appears to be the only potential adverse effect on applicants.

6.2.4.2.8 Compliance with Legal Issues and EEOC Guidelines

There has been some research done to determine whether the GPP-I discriminates among racial and ethnic groups. Hays (cited in Ref. 11) conducted a study using fifty-eight black, forty-eight Hispanic, and 771 white student teachers as subjects. All subjects were administered the GPP-I upon graduation, just prior to beginning their initial teaching assignments. Significant differences among groups were found on two scales: Blacks scored higher than whites on measures of Cautiousness and Emotional Stability. There were no significant differences in scores between the Hispanic subjects and either the white or black subjects.

In a study by Gordon (1973, cited in Ref. 11), the GPP-I was administered to a total of 181 female service (food, laundry, maintenance, etc.) personnel in 3 mental hospitals in Pennsylvania. The sample consisted of 102 black and 79 white subjects. Results showed significant differences on two scales: whites scored higher on Responsibility, while blacks scored higher on Original Thinking. The author noted that the results may have been partially due to group differences in age and education, the white subjects being older and having had less formal education than the black subjects.

Gordon (Ref. 11) noted that overall, the significant racial and ethnic differences demonstrated by the research with respect to the GPP-I, were both small in magnitude and in directions that would be unlikely to result in adverse impact to minority group members for selection purposes.

Since the GPP-I has been shown to be unlikely to discriminate against minority group members, this instrument appears to be basically in accordance with EEOC Guidelines. This review did not reveal any legal cases which centered upon the use of the GPP-I.

6.2.4.2.9 Confidentiality of GPP-I Results

Because the GPP-I is a psychological test, it is subject to the same ethical standards for confidentiality as previously discussed for such instruments.

6.2.4.2.10 Susceptibility To Faking

Gordon (Ref. 11) stated that although the GPP-I does not have any "lie" or "faking" scales, the construction of the instrument is such that individuals can choose an excess of complimentary alternatives without detracting from the reliability of the test. This issue has been investigated in several studies discussed by Gordon (Ref. 11), who concluded that although some distortion may occur in situations where individuals are motivated to provide favorable responses, the magnitude of this distortion appears to be relatively small. The

author stated that overall findings indicate that individuals, on the average, tend to provide two extra complimentary responses in situations where they are asked to "fake good" as compared to normal test-taking situations.

6.2.4.2.11 Labor Relations Considerations

This review did not reveal any labor relations controversies which centered on the use of the GPP-I, although the possibility of such action could occur if it were determined that this instrument violated legal or ethical guidelines.

6.2.4.2.12 Reevaluation Considerations

In accordance with EEOC guidelines, a job applicant screened out on the basis of the GPP-I should be provided with an opportunity for reevaluation. As with most personality tests, this option appears especially important due to the inclusion of some traits on the instrument which might fluctuate due to situational changes. Such changes could result from real-life events (e.g., marital or financial problems) that might temporarily alter certain aspects of an individual's personality.

6.2.4.2.13 Applicability of the GPP-I to Selection of Nuclear Facility Personnel with Regard to Emotional Instability

The GPP-I includes a scale specifically designated as an index of "Emotional Stability". The inclusion of this scale, as well as the occupational normative data provided in the test manual, appear to render this test worthy of consideration as a selection instrument for nuclear facility personnel.

One apparent weakness of the GPP-I is its lack of a scale included to detect faking of responses. Although, as Gordon (Ref. 11) pointed out, the structure of the instrument appears to minimize the potential for distortion, this deficiency must be kept in mind, especially in considering the use of this instrument in a personnel selection procedure.

As with the other personality tests discussed, validation of the GPP-I on nuclear facility personnel would appear a logical step to its consideration for inclusion as a selection instrument.

6.2.5 Eysenck Personality Questionnaire (EPQ)

6.2.5.1 Overview

The Eysenck Personality Questionnaire (EPQ) is a self-report inventory of the "Yes-No" (a variation of "True-False") format. It was published in 1975, and represents the most recent revision of a series of personality measures developed

by Eysenck and his colleagues. These earlier instruments included the Maudsley Medical Questionnaire (Cited in Ref. 15), the Maudsley Personality Inventory (Ref. 15), and the Eysenck Personality Inventory (Ref. 13).

The EPQ consists of ninety items which yield scores on the following three personality dimensions: Extraversion-Introversion (E), Neuroticism (N) and Psychoticism (P). In addition, there is a Lie Scale (L) designed to detect "faking good" response sets. The authors of the test suggest that in certain contexts, and for practical purposes, the Neuroticism Scale be referred to as "emotionality" or "stability-instability", and the Psychoticism Scale as "tough-mindedness" (Ref. 16, p.3). They view the E and N traits measured by the EPQ as contributing more to a description of an individual's personality than any other two factors excluding those that pertain to the cognitive or intellectual domain. The inclusion of the P scale is the major difference between the EPQ and its predecessor, the EPI. Eysenck and Eysenck (Ref. 16) view the term "psychoticism" to refer to a trait which they suggest is present in everyone to some degree. They stress that a high score on this scale represents a predisposition to the development of a psychiatric abnormality but scores on the scale do not necessarily predict the development of psychiatric abnormality.

As measured by the EPQ, the typical extravert is described as tending toward being aggressive, losing his temper easily, displaying impulsivity and unreliability, and being carefree and easy-going. The typical introvert is viewed as being quiet and retiring, reserved and somewhat uncomfortable with others, serious-minded, well-organized, and rather pessimistic. The higher N scorer is described as an anxious worrier who is frequently depressed, easily upset and overly reactive, somewhat rigid, prone to psychosomatic disorders and bearing a constant preoccupation with things that might go wrong. A high P scorer is seen as being a loner, being uncaring and often cruel to others, lacking in empathy and interpersonal sensitivity, displaying hostile or aggressive actions with little regard for others, and exhibiting a need for new sources of stimulation, with little regard for danger (Ref. 16).

The normative data for the EPQ was gathered on a total of 5,574 "normal" subjects and 2,154 "abnormal" subjects (the term "abnormal" is used here to refer to psychiatric patients, prisoners, drug addicts, and alcoholics who participated in the standardization studies). Eysenck and Eysenck (Ref. 16) pointed out in the Manual for the EPQ, that the "normal" subjects who provided the standardization data were largely residents of urban areas, and displayed large sex and age differences. In order to compensate for these effects, the standardization tables are broken down according to age groups and sex. The Manual also includes tables depicting

the means and standard deviations for a variety of occupational groups.

The EPQ is available in one form only for adult populations. It is quickly and easily scored by hand, thus machine-scoring services are not available. Rather than being converted to Standard Scores, raw scores on the EPQ scales are compared directly to the means and standard deviations for the appropriate normative data tables.

6.2.5.2 Relevant Considerations

6.2.5.2.1 Reliability

With respect to the EPQ, traditional evaluation studies have focused on the test-retest and internal consistency reliability approaches to measuring the reliability of this instrument.

Eysenck and Eysenck, (Ref. 16) cited several studies designed to evaluate the test-retest reliability of the EPQ. The samples used for these studies included 111 dental students, 31 polytechnic students, 40 social workers, and 55 university students. All subjects were retested following one-month intervals. Reliability coefficients for all groups ranged from .51 to .96, with an overall mean of .84.

Eysenck and Eysenck (Ref. 16) reported a study designed to measure the internal consistency reliability of the EPQ, using 500 male and 500 female adult subjects. Alpha coefficients (a measure of internal consistency reliability) ranged from .68 to .85 for both groups, with an overall mean of .80. There were no significant differences based on sex of the subjects. A similar study (Ref. 16) using a sample of prison inmates revealed coefficient alpha values comparable to those in the investigation discussed above, with an overall mean of .82.

6.2.5.2.2 Validity

With respect to the EPQ, traditional evaluation studies have centered on the construct validation approach to the measurement of validity.

In discussing measures of construct validity for the EPQ, Eysenck and Eysenck (Ref. 16) reviewed studies comparing the scales on this instrument to those of various other personality tests. Correlations between the EPI and Cattell's Anxiety Factor (Ref. 17) and Neuroticism Factor (Ref. 18) ranged from .34 to .81, with a mean of .59. Comparisons of the EPQ and the Multiple Affect Adjective Checklist (Ref. 19) revealed validity coefficients ranging from absolute values of .00 to .59, with a mean of .25. Another study compared the EPQ to the California Psychological Inventory (Ref. 20).

This study revealed validity coefficients ranging from .01 to .67, with a mean of .29.

6.2.5.2.3 Cost of Development and Administration

The current prices of the materials necessary for administration of the EPQ are listed below:

EPQ Manual -	$ 2.00
Package of 500 EPQ Forms -	57.00
Sets of Handscoring Keys for P, E, N and L Scales of EPQ (per set) -	2.75

6.2.5.2.4 Administrator: Training and Qualifications

No specific training or education is required for administration of the EPQ, since it is a self-report inventory. The administrator should, however, be provided with the basic instructions to give the test-takers. Interpretation of EPQ scores should be conducted by a qualified psychiatrist or psychologist who is thoroughly familiar with the instrument.

6.2.5.2.5 Administration of EPQ: Time and Difficulty

The EPQ is a relatively brief personality inventory, and can usually be completed in fifteen to twenty minutes. There are no apparent difficulties associated with test administration.

6.2.5.2.6 Equipment/Materials Needed for Administration

For pre-employment screening purposes, the only materials needed for administration of the EPQ, are the Adult EPQ Forms and pens or pencils.

6.2.5.2.7 Personal Effects on Applicants

Although the authors of the EPQ (Ref. 16) state that the test items are not socially objectionable, there are no data available to support this contention. Some individuals might be likely to view some of the items as offensive or invasions of privacy, however, this appears to be the only apparent potentially detrimental effect to the applicant.

6.2.5.2.8 Compliance with Legal Issues and EEOC Guidelines

Although the manual provides separate standardization tables based on age, sex, and various occupational groups, it must be stressed that the EPQ was standardized on a primarily urban population. For this reason, there are some questions regarding the administration of this test to non-urban populations, and the potential discriminatory effects.

This possibility could result in conflicts with EEOC guidelines if it results in adverse impact in an employment setting.

6.2.5.2.9 Confidentiality of EPQ Results

Because the EPQ is a psychological test, it is subject to the same Ethical Standards for confidentiality as previously discussed for such instruments.

6.2.5.2.10 Susceptibility to Faking

The Lie scale (L) on the EPQ was developed to aid in the detection of response distortion on this instrument, and was validated along with the other scales. The trained and qualified interpreter should, therefore, be able to detect response sets suggestive of "faking good."

6.2.5.2.11 Labor Relations Considerations

This review did not reveal any labor relations controversies which centered on the use of the EPQ. The possibility of such action cannot be ruled out, however, if it were determined that this instrument violated legal or ethical guidelines.

6.2.5.2.12 Reevaluation Considerations

In accordance with EEOC Guidelines, a job applicant screened out on the basis of the EPQ should be provided with an opportunity for reevaluation. As with most personality tests, this option appears especially important due to the inclusion of some items or traits on the test which might reflect real-life situational changes in an individual at different testing times.

6.2.5.2.13 Applicability of the EPQ to Selection of Nuclear Facility Personnel with Regard to Emotional Instability

As stated by the authors of the EPQ (Ref. 16), the Neuroticism Scale may also be considered a measure of an individual's degree of emotional stability/instability. For this reason, it appears relevant to the purposes of this review.

The major weakness of the EPQ appears to be the fact that its normative data were obtained primarily from urban dwellers. Since applicants to nuclear power facilities typically come from all geographic regions, it would appear advisable to gather additional normative data from residents of suburban, small-town, and sparsely populated communities, prior to considering the use of this instrument.

6.3 Physiological Measurement Techniques

6.3.1 Overview

Physiological measurement techniques may be defined as those methods designed to measure internal bodily functions. This definition is necessarily broad in nature due to the complexity of the human body and the vast number of specific parameters that might be measured. For the purposes of this review, this section will focus on techniques designed to detect physiological arousal patterns occurring in response to environmental stressors. The rationale underlying the consideration of physiological measurement techniques in the selection of nuclear facility personnel represents the interactional pattern of the behavioral, emotional, and biological aspects of human functioning, and their combined impact on both physical and mental health.

There are a variety of typical physiological changes which may occur in an individual confronted with a stressor. Examples of these bodily responses include increased skeletal muscle tension, alterations in brain wave activity, increased heart rate, respiration rate and blood pressure, fluctuations in skin conductance levels, and a variety of hormonal and neurochemical imbalances (Refs. 21 and 22). If these changes occur repeatedly, they predispose the person to the development of psychophysiological/psychosomatic disorders such as hypertension, migraine headaches, asthma, gastro-intestinal disorders, etc.

Despite the widespread use of polygraph testing for lie-detection purposes in employment settings, there have been few, if any, attempts to employ physiological evaluation of stress responding in personnel selection procedures. Before such an approach could be considered for implementation, it would be necessary to establish significant criterion-related validity for physiological measurement techniques. In order to justify its use, these data would have to indicate a strong relationship between the predictors (the physiological measurement techniques selected for use) and the criterion (performance on the job).

Several early investigations (Refs. 23 and 24) established a theoretical foundation relating degree of stress to level of performance. These studies led to the formulation of the Yerkes-Dodson Law (Ref. 24), more commonly known as the "inverted-U" relationship between arousal levels and performance. This hypothesis postulates an inverted U-shaped graphic representation of this interrelationship, indicating that there is a moderate level of stress or arousal associated with maximum performance, such that either too little or too much stress induces performance decrements. Easterbrook (cited in Ref. 25) suggested an explanation for this phenomenon. He pointed out that when levels of arousal are too low,

individuals tend to become less vigilant and ignore relevant cues in the environment, thus hampering effective performance. As arousal increases, so does selective attention, so that relevant cues are dealt with, and irrelevant cues are ignored. If the level of stress becomes too high, the simultaneously increasing degree of selectivity leads to the screening out of relevant information, so that once again, a decrease in performance is seen. Sjoberg (Ref. 25) conducted a study designed to determine the effects of varying task difficulty on this working hypothesis. He found that the optimal arousal level for effective performance varied indirectly with the level of task difficulty. More specifically, the more difficult or demanding a task, the lower the arousal level needed for optimal performance. Conversely, easier tasks are best performed when accompanied by higher levels of physiological arousal.

There has been some research (Refs. 26, 27, and 28) conducted relating physiological stress profiles to performance in occupational settings, using subjects already employed, in attempts to provide empirical data relating to the early theoretical foundations. The results of these investigations provided inconsistent data which are at least partially attributable to wide variations in methodology. These findings illustrate the need for both precise specification of performance evaluation criteria and clear delineation of acceptable physiological profiles of stress responding.

There is a wide range of physiological functions which may be measured, and a number of specific instruments designed for measurement purposes. For the purposes of this review, the instruments to be discussed are included on the basis of their applicability and actual usage in the measurement of stress responding (a topic which subsumes physiological detection of deception). It is recognized that there are some techniques which are available for physiological measurement but are typically used only for research purposes, such as biochemical analysis and measurement of pupillary changes. Such research techniques will not be discussed here.

6.3.2 Biofeedback Instruments

6.3.2.1. Overview

Within approximately the past two decades, numerous researchers with various types of training* (Refs. 29, 30, 31, 32, 33, and 34) have compiled vast amounts of literature indicating that what were previously considered strictly involuntary physiological mechanisms (e.g., heart rate, blood pressure, galvanic skin response, brain-wave activity, etc.) can be brought under voluntary control, implying a mind-body interaction. This preliminary research evolved into the area of biofeedback research, a rapidly expanding field which encompasses a vast range of current and potential applications. A basic working definition of the term biofeedback, which is applicable to all of the specific areas of investigation was provided by Blanchard and Epstein (Ref. 35) who described biofeedback as "a process in which a person learns to reliably influence physiological responses of two kinds: either responses which are not ordinarily under voluntary control or responses which ordinarily are easily regulated, but for which regulation has broken down due to trauma or disease" (p.2). This learning typically requires the use of some variety of instrumentation, designed to provide the individual with feedback regarding the functioning of the physiological system being monitored, so that learning can take place.

The term biofeedback implies that such instruments possess capabilities of feeding back information to the individual being monitored. The purpose of this feedback in traditional clinical usage has been to provide the person with objective information regarding his/her physiological responses, so that he/she can learn to relate this to subjective feeling states and thus gain voluntary control over the physiological functions being measured. Such equipment can be operated in a feedback-off mode, however, so that only the equipment operator is in the position of detecting physiological measurements. Thus, the data obtained can be interpreted without the possible confounding factor of learning taking place

* J.T. Hart, "Autocontrol of EEG Alpha, "Paper presented at the annual meeting of the Society for Psychophysiological Research, San Diego, California, October (1967).

in the actual measurement situation. The capability of feedback must be considered as a potential advantage, however, to individuals or organizations who might, at some point, become interested in implementing stress management programs.

The vast collection of data resulting from the studies investigating available biofeedback techniques and future possibilities, suggests that despite some current limitations, biofeedback represents a totally new and promising way of gathering previously unattainable data, teaching people to control their own physiological functions, treating certain disease states and measuring physiological responses to stress. It also appears to be a viable way of relating objective physiological data regarding a person's physiological functioning to his/her self-report indices of subjective feeling states, as well as, to his/her personality traits, general behavioral tendencies, and job performance. The following list provides brief descriptions of the specific types of biofeedback instruments used in the measurement of stress responses.

Instrument	Description
• Electromyograph (EMG)	Provides measure of electrical activity in motor neurons, which may be interpreted as index of skeletal muscle tension; when frontalis (forehead) muscle is monitored, provides index of general muscle tension from top of head to bottom of rib cage, generally interpreted as an index of overall body muscle tension.
• Peripheral Skin Temperature Feedback Instrumentation	Detects changes in amount of blood flow to extremities (typically fingertip) through use of temperature-sensitive thermistor providing reading in degrees Fahrenheit; amount of peripheral blood flow and resulting skin temperature reflect degree of vasoconstriction of peripheral blood vessels, presumed to be an index of sympathetic nervous system arousal or stress reactivity.
• Pulse Rate Feedback Instrumentation	Measures heart rate either instantaneously or by averaged reading of fixed amount of time or fixed number of beats.

• Galvanic Skin Response (GSR)	Detects changes in electrical resistance of the skin resulting from changes in arousal; signal has two characteristics: 1) a slowly changing (tonic) level, indicative of general arousal level, and 2) shorter, more abrupt changes (phasic) presumed to reflect an immediate response to a stimulus just presented.
• Electroencephalograph (EEG)	Measures the electrical activity of various parts of the brain in terms of frequency (Hz.) and amplitude of waveforms; there are four main patterns of brainwaves, determined by frequency range; the alpha range is generally considered to be from 8-12 Hz., and is representative of relaxed wakefulness in the occipital cortex region of the brain; Beta waves, those with frequencies above 13 Hz. are interpreted as signs of arousal, problem-solving, or stress-reactivity, depending upon the situation.

6.3.2.2. Relevant Considerations

6.3.2.2.1 Reliability

With respect to biofeedback instrumentation, this review did not reveal any literature pertaining to the traditional forms of reliability measurement. This may be the case because biofeedback equipment is essentially a form of electronic instrumentation. Thus, the accuracy or reliability of biofeedback equipment would probably be more appropriately evaluated in terms of the quality of the instrumentation and the absence of artifacts (i.e., conditions that provide inaccurate readings). The quality of the instrument varies directly with its sensitivity, thus, high-grade equipment would appear to be more reliable than instruments of inferior quality.

Thorough training of equipment operators is necessary for the prevention of artifacts. Factors that could produce false readings include bodily movement of the subject, improper electrode placement, movement of electrode cases and proximity to other electrical equipment.

6.3.2.2.2 Validity

As with reliability, this review did not reveal any literature pertaining to the traditional validation approaches with respect to biofeedback devices. With respect to this review, it appears that research efforts aimed at predicting performance and stress reactivity on the basis of biofeedback readings, would be a beneficial pursuit. The optimal setting for conducting such research would be in an actual nuclear facility. This type of approach would serve as an effort to establish criterion-related validity for biofeedback instrumentation.

6.3.2.2.3 Cost of Development and Administration

There are at least a dozen manufacturers of high quality biofeedback equipment, and numerous manufacturers of lesser grade devices. An initial consideration in regard to the selection of instruments would involve determination of which physiological functions would provide the most useful information. A high quality instrument, having wide enough ranges of measurement to be optimally sensitive and accurate, could cost anywhere from several hundred to two thousand dollars. Thus, procurement of various manufacturers' specifications is a logical first step for any organization considering the use of such equipment.

6.3.2.2.4 Administrator: Training and Qualifications

An individual can usually be trained in the operation of a specific parameter biofeedback instrument within approximately one week. Some instruments (e.g., EEG and GSR) are more complex than others for operational purposes. Interpretation of the data obtained from biofeedback instruments requires considerable understanding of the physiology involved. Specific qualifications for biofeedback training professionals exist in some states and are variable from state to state. These requirements serve the purpose of certifying individuals who will be using such equipment for training, however it is likely that the same requirements would also apply to persons using the equipment for monitoring purposes only.

6.3.2.2.5 Administration of Measurement Technique: Time and Difficulty

A typical length of time for one biofeedback session is about twenty minutes; this duration is appropriate for establishing a stable baseline level of the parameter(s) being measured.

Although all the major instruments now in use rely on skin surface electrodes (as opposed to needle electrodes) this implies both proper skin preparation and the need to make sure that the electrodes are placed on the identical body site for repeated measurements.

6.3.2.2.6 Equipment/Materials Needed for Administration

Most biofeedback instruments are modular, self-contained, and are typically battery-powered, eliminating the possibility of electrical shock to the individual as well as the inaccuracies of recording posed by occasional power surges. Many manufacturers now produce modular data integrator equipment which can receive inputs from all the instruments being used to monitor the individual. Such units can be programmed to collect data on all functions at precise time intervals, a task virtually impossible for human recording. In addition, some of these integrators will perform simple statistical analyses of the data as they are being recorded.

The basic equipment needed for utilizing biofeedback instruments to measure stress responding includes the following items:

(1) modular feedback instrument for each parameter being measured

(2) skin surface electrodes

(3) electrode cables

(4) alcohol and cotton balls for preparation of body site

(5) data acquisition and integrator component, if desired, with cables to be attached to each individual modular component

(6) data recording sheets

6.3.2.2.7 Personal Effects on Applicants

It is possible that some individuals might view physiological measurement as an invasion of privacy and thus object to this procedure. There are also occasional cases in which subjects are fearful of being electrically shocked by the instrumentation. This problem may be resolved by having the administrator explain to the subject that the current passes only from the person to the machine, and not in the reverse direction.

6.3.2.2.8 Compliance with Legal Issues and EEOC Guidelines.

Because biofeedback represents a unique combination medical/psychological measurement technique and has not been used significantly as a selection device, it is difficult to determine whether it would be subject to the same EEOC guidelines as other psychological tests. Biofeedback is a relatively new field of endeavor, thus its reliability and validity have not yet been adequately established. This review, while revealing no evidence suggestive of adverse impact, does not necessarily indicate compliance of this procedure with established guidelines.

6.3.2.2.9 Confidentiality of Measurement Technique Results

Because data gathered through the use of biofeedback instruments are personal in nature, they should be treated with the same confidentiality considerations as psychological or medical data. That is, they should be released only to persons who are clearly in positions of needing this information for personnel selection decisions.

6.3.2.2.10 Susceptibility to Faking

It is unlikely that an individual applying for a nuclear facility position would attempt to modify this physiological functioning while being measured by biofeedback equipment, since to do so would increase arousal levels and thus potentially decrease his/her suitability for the job. It would be possible to distort such a profile by such actions as clenching teeth, tensing muscles, wiggling toes, etc. These activities would produce a profile reflecting consistently high arousal levels, so that discrimination of significant physiological reactions to stimuli presented would be difficult. The administrator should carefully observe the individual being monitored in an effort to detect such attempts at distortion.

6.3.2.2.11 Labor Relations Considerations

This review revealed no labor relations controversies which centered on the use of biofeedback techniques within a selection context. The possibility of such an occurrence cannot be ruled out, however, if such use were to be shown to violate legal or ethical guidelines.

6.3.2.2.12 Reevaluation Considerations

In accordance with EEOC Guidelines, an individual screened out on a basis of biofeedback results should be given the opportunity for reevaluation. The stress response is likely to fluctuate in a fashion similar to that of certain situational personality variables, providing further justification for reevaluation.

6.3.2.3 Applicability of Biofeedback Instruments to Selection of Nuclear Facility Personnel with Regard to Emotional Instability

It appears that biofeedback techniques have the potential for providing significant additional information useful in the selection of nuclear facility personnel. The optional use of these techniques would involve the measurement of physiological indices during situational simulations and comparison of these data to baseline rates. The major drawback to this application appears to be the potential artifacts produced by body movement.

Another reservation regarding biofeedback is the lack of established validity relating stress levels to on-the-job performance. This appears to be a fruitful area for further investigation.

6.3.3 Polygraphs

6.3.3.1 Overview

Polygraphic equipment was originally developed for the purpose of research investigations regarding physiological functioning. The complexity of the original instruments was due to the incorporation of mechanisms designed to compensate for poor quality of amplified signals resulting from inadequate electrode techniques, unreliability of amplifiers, and the need for complete electronic shielding (Ref. 36). Improvements in modern polygraphs have greatly reduced such sources of error; however, polygraphic instruments remain quite complex due to their capabilities of simultaneously monitoring a vast range of physiological functions, thus implying the use of an equally broad span of accessory paraphernalia.

The two current major applications of the polygraph include physiological research and detection of deception. The application of polygraphy to detection of stress responses (other than responses to deception) remains primarily a research endeavor. This is due to the complexity of laboratory polygraph equipment operation and the development of modular biofeedback equipment which can provide the same data with relatively greater ease of operation. For these reasons, this section will review the technique of polygraphy primarily from the viewpoint of its major applied use, detection of deception. This topic is relevant to the objectives of this review because deception is frequently an index of emotional instability (see Section 3.2.1.2).

The type of polygraphic equipment generally used for lie-detection purposes is known as the field polygraph. Field polygraphs are portable, modified versions of standard laboratory equipment, which are usually designed to measure respiration rate, blood pressure, heart rate, and skin conductance levels (Ref. 37).

Barland and Raskin (Ref. 38) outlined the three basic types of approaches used in field polygraphic lie detection. The first method is known as the "peak of tension" technique. With this approach the applicant, having had the polygraphic sensing devices attached to his/her body, is asked a series of questions. This series contains one critical item and several non-critical items which are in some way similar to the critical items. The questions are presented in a standard sequence, which the applicant knows ahead of time. If the individual's highest levels of physiological arousal are demonstrated in response to the critical item, he/she is

judged deceptive. If his/her highest response levels occur with a different item, he/she is judged truthful. The second approach, known as the "relevant-irrelevant" technique is conducted in a fashion similar to the peak of tension method; however, in this case, the applicant is unaware of the sequence of the questions, more than one critical item is included, and the critical items are not similar to the non-critical items. The third approach, known as the "control question" technique was developed by Reid (Ref. 39) in an effort to provide a more accurate means of detecting deception in individuals with chronic high or low arousal levels. In this situation, both the relevant and the irrelevant (control) items are designed to elicit high arousal levels in applicants. The theoretical basis for this approach involves the notion that truthful individuals will display higher levels of reactivity to control questions than to critical items, while deceptive persons will exhibit higher arousal levels in response to critical or relevant questions.

The use of the polygraph as a lie-detection device is, and has been, a controversial issue since the beginning of this application of the instrument some fifty years ago. This ongoing dispute involves a variety of issues. Recent estimates indicate that between two-hundred thousand to a half million lie-detector tests are administered annually in the private sector (Ref. 40). In their study of major U.S. corporations in regard to polygraph usage, Belt and Holden (Ref. 40) found that the three major purposes cited by corporations using the polygraph were to verify employment applications, to conduct periodic evaluations of employee honesty, loyalty and adherence to company policy, and to investigate isolated incidents of theft, vandalism, sabotage, etc. The proponents of the use of polygraphy for lie-detection cite the low operating costs ($25.00 to $50.00 per test), speed of administration, and advantages of additional information useful in selecting out individuals who would be potential company liabilities. Those who oppose the use of this technique view the test as an invasion of privacy, as well as stating that other selection methods available are adequate. (Ref. 40).

The two major issues surrounding the use of lie-detection by polygraphy involve the legalities of this technique and the questions regarding its validity. These two areas of concern will be addressed under Relevant Considerations.

6.3.3.2 Relevant Considerations

6.3.3.2.1 Reliability

With respect to the polygraph, traditional evaluation studies have centered on the test-retest and inter-rater approaches to the measurement of reliability. Caution should be exercised in interpreting the results of these studies, since quite different findings have been reported by proponents vs. opponents of the use of this technique.

As noted by several authors (Refs. 41 and 37) research regarding the reliability of the use of the polygraph for lie detection has remained generally inconclusive due to lack of standardized research design.

Balloun and Holmes (Ref. 41) conducted a study designed to examine the effects of repeated polygraphic examinations on ability to detect deception. Their results demonstrated greatly diminished lie detection capabilities upon repeated testing.

Studies investigating the inter-rater reliability of the polygraph for lie detection purposes (Refs. 38, 42, 37, 43, and 44) have generally compared multiple raters' determinations of truthfulness or deception, based on visual inspection of individuals' polygraph charts. The findings from these investigations exhibit generally high percentages of inter-rater agreement, with a mean of approximately 88%. These high inter-rater agreement percentages must be viewed with caution, however, because agreement among raters does not ensure that the polygraph is accurate in detecting deception or truthfulness (see section on Validity which follows).

6.3.3.2.2 Validity

With respect to the polygraph, traditional evaluation studies have centered on the construct validation approach to the measurement of validity.

Studies designed to measure the validity of the polygraph for lie detection applications have traditionally compared conclusions derived from polygraphic charts with other external measures of deception or truthfulness. These external measures have included sources of information such as background checks, confessions of guilt in other situations, and court records of conviction. Validity data obtained from studies employing these types of comparisons have revealed large variations, ranging from chance (50%) or less accuracy (Ref. 45) to figures as high as 81% (Ref. 38), and 88% (Ref. 43).

It should be noted that it is difficult to compare the available research findings regarding the validity of polygraphic lie detection techniques due to extreme variations in methodological approach.

These variations in approach include lack of uniformity in the following kinds of research design considerations: types and combinations of questions asked, physiological parameters measured, field vs. laboratory interpretation of response patterns, number of raters evaluating each chart and differing criteria for determination of deception. These methodological problems should be considered in the interpretation of studies designed to measure the validity of the polygraph for use in detection of deception.

6.3.3.2.3 Cost of Development and Administration

There are several manufacturers of polygraphic equipment, each having their own set of instrumentation prices. Most polygraphs are quite expensive, with costs for complete recording systems ranging from around $2,000 to $10,000. The approximate cost of a particular system depends not only on the quality and sophistication of the instruments desired, but also on the number of physiological parameters to be measured. In practice, most polygraphic lie detection procedures do not require more than a four-channel acquisition system, capable of monitoring four physiological functions simultaneously.

Due to the expense and expertise necessary for operation of polygraphic equipment, many organizations send job applicants to private polygraph examiners, who generally charge from $25.00 to $50.00 per test.

6.3.3.2.4 Administrator: Training and Qualifications

As noted by Belt and Holden (Ref. 40), nineteen states* thus far have passed formal laws outlining standards for the training and licensing of polygraph examiners. The remaining states do not as yet require any formal certification for polygraph administration. Thorough training of test administrators in the areas of instrumentation operation, physiological chart interpretation and methods of questioning would appear, however, to be the minimal requirements necessary for effective polygraph administration.

6.3.3.2.5 Administration: Time and Difficulty

A polygraphic lie detection test typically requires from a half-hour to forty-five minutes for administration. The specific time required depends on both the number of questions asked and the type of approach used by the polygraph examiner.

There are no inherent difficulties associated with administration of a polygraph test; however some potential problems

* These states include Alabama, Arizona, Arkansas, Florida, Georgia, Illinois, Kentucky, Michigan, Mississippi, Nevada, New Mexico, North Carolina, North Dakota, Oklahoma, South Carolina, Texas, Utah, Vermont, and Virginia.

include poor electrode contact, proximity of other electronic equipment and applicants possibly not understanding some of the questions asked.

6.3.3.2.6 Equipment/Materials Needed for Administration

The basic equipment necessary for polygraph test administration includes the following instruments and accessories:

(1) Polygraphic acquisition instrument

(2) Input couplers and transducers for each physiological function measured

(3) Skin surface electrodes

(4) Electrode cables and ground wires

(5) Recording paper and ink

As with biofeedback equipment, most polygraphic manufacturers also produce special feature accessory equipment, such as data integration instruments; however, such adjunct devices are not typically employed in lie detection applications of the polygraph.

6.3.3.2.7 Personal Effects on Applicants

Belt and Holden (Ref. 40) noted that the most significant objection maintained by opponents of polygraphic lie detection tests is this technique's invasion of privacy. It seems reasonable that any applicant required to undergo a lie detection test might experience feelings of personal intrusion. This is due to the very nature of the technique in its implication that the information given by the applicant requires external verification before it will be accepted as being true.

Another important potential personal effect on applicants involves the ramifications of the results of the test. If an applicant were not selected on the basis of information revealed by a lie detection test, there could be significant emotional reactions incurred by the individual, especially in a case where the test results were inaccurate in depicting a person as deceptive.

6.3.3.2.8 Compliance with Legal Issues and EEOC Guidelines.

In their discussion of the legalities regarding the use of the polygraphic lie detector test, Belt and Holden (Ref. 41) pointed out that individual state governments have begun to enact legislation aimed at regulating the use of the lie detector test. In addition to the nineteen states which have enacted legislation governing the training and licensing of

polygraph examiners, fifteen other states* now have laws which to some degree place restrictions on the use of the polygraph in employment practices. The specific provisions of these laws vary from state to state, ranging from some that prohibit businesses from requiring employees to submit to lie detector tests, to those of a further extreme of banning employers from even asking their personnel to undergo polygraph exams.

The existing legislation regarding polygraph test usage in employment stems not only from ethical concerns such as violations of rights to privacy, but also, and probably equally as important, from the continued controversy regarding the validity of such procedures. Belt and Holden (Ref. 40) noted that as a result of these concerns, active positions against the use of the polygraph for lie detection have been adopted by the American Civil Liberties Union (ACLU), the Federal Privacy Protection Study Commission, and several Congressmen. There is also a possibility that polygraph examiners might discriminate among the applicants they test on the basis of gender or other demographic characteristics. This possibility plus the lack of adequately established validity renders this technique questionable in terms of compliance with EEOC guidelines.

6.3.3.2.9 Confidentiality of Measurement Technique Results

As with other personal information regarding job applicants revealed by personnel selection procedures, determinations of truthfulness or deception based on polygraph tests should be kept confidential. Only those persons who are clearly involved in the selection process or in subsequent personnel matters within an organization should have access to polygraph test results.

6.3.3.2.10 Susceptibility to Faking

Although determinations of truthfulness or deception on the basis of polygraph tests are made by examiners' evaluations of the charts, the instruments measure only patterns of physiological arousal, which are then open to human interpretation. Therefore, as in the case of biofeedback equipment, it

* These states Alaska, California, Connecticut, Delaware, Hawaii, Idaho, Maryland, Massachusetts, Minnesota, Montana, New Jersey, Oregon, Pennsylvania, Rhode Island, and Washington.

is possible for an individual being monitored to intentionally distort his/her polygraphic results by engaging in some activity (such as biting his/her tongue or breathing rapidly) which will serve to produce high levels of physiological arousal in response to control questions. As pointed out by Lykken (Ref. 42) strong responses to control questions make it difficult to interpret responses to relevant or critical items.

Although most qualified polygraph examiners routinely check for attempts at distortion during testing, it is doubtful that all such actions can be easily detected. For these reasons, the possibility of applicants successfully faking polygraph lie detection tests cannot be ruled out.

6.3.3.2.11 Labor Relations Considerations

Belt and Holden (Ref. 40) noted that the American Federation of Labor-Congress of Industrial Organizations (AFL-CIO) has taken an active stance against the use of the polygraph for lie detection in employment settings. This position results from both this technique's alleged invasion of privacy as well as grievances claimed by employees and job applicants who maintain that they have been wrongly judged deceptive by such tests. Thus, it appears that labor disputes involving the use or the polygraph are likely to occur.

6.3.3.2.12 Reevaluation Considerations

In accordance with EEOC guidelines, an individual screened out on the basis of a polygraph test should be given the opportunity for reevaluation. If the polygraphic lie detection technique were a proven valid procedure for detecting deception, it is unlikely that the results to the same set of questions would be different for the same applicant upon reevaluation. However, as noted by Balloun and Holmes (Ref. 41), this technique's effectiveness in detecting deception decreases with repeated administrations, so that reevaluation results might have to be considered invalid.

6.3.3.2.13 Applicability of the Polygraph to Selection of Nuclear Facility Personnel with Regard to Emotional Instability

It appears that the use of the polygraph for detection of deception as part of a selection procedure, cannot, at this point, be given substantial support. The questionable reliability and validity of this technique, as well as the issues involving its legality, prevent its recommendation as part of a screening procedure for emotional instability with nuclear facility personnel applicants.

6.3.4 Telemetry

6.3.4.1 Overview

Telemetry refers to a technique encompassing a variety of specific devices which were originally designed for use with astronauts, as a means of obtaining prolonged physiological measurements without the use of cables attached to equipment (Ref. 46). The original device employed in space flight utilized a silastic-coated gelatin capsule designed to be swallowed or inserted vaginally or rectally. The capsule contained electronic sensors to transmit recordings of deep body temperature and heart rate to a receiving antenna built into a belt worn around the individual's waist. The receiving antenna in the belt was attached to a transmitter designed to relay the incoming data to receiver equipment as a designated data collection center on earth.

Unfortunately, there has been relatively little subsequent research investigating telemetry capabilities. One major development in the use of telemetry techniques for application on earth, however, has been the invention of transducers which are applied in a similar fashion as skin surface electrodes, thus eliminating the necessity for ingestion of any type of device. These transducers are capable of monitoring such physiological functions as skeletal muscle tension and brain-wave activity, and are now commercially available through a few manufacturers. The greatest potential advantage of this type of device over other physiological measurement techniques is the lack of cables connecting the individual to the monitoring equipment. This factor provides for freedom of movement and eliminates the artifacts produced by movement of individuals who are attached to physiological measurement equipment by cables. Such a consideration would be quite relevant to the measurement of arousal levels in response to simulated stress.

6.3.4.2 Relevant Considerations

6.3.4.2.1 Reliability

With respect to telemetry techniques, this review did not reveal any research studies which have attempted to measure the reliability of telemetry instrumentation.

6.3.4.2.2 Validity

With respect to telemetry techniques, this review did not reveal any completed studies which have attempted to measure the validity of these procedures.

6.3.4.2.3 Cost of Development and Administration

Telemetry devices are currently available through a few of the older and well-established manufacturers of physiological measurement instrumentation. The general price range is approximately $1,700 to $2,000 for a complete system, with specific costs varying by company.

6.3.4.2.4 Administrator: Training and Qualifications

Since telemetry is still used on a very limited basis, there are at present no laws governing the training and qualifications of administrators. It would seem, however, that a strong physiology background and thorough knowledge of instrument operation would be necessary prerequisites to effective administration.

6.3.4.2.5 Administration of Measurement Technique: Time and Difficulty

There is no fixed time for administration of telemetry techniques. The duration of physiological measurement sessions should, however, be of sufficient length (approximately twenty to thirty minutes) to permit the administrator to obtain discernible patterns of physiological functioning.

There could be some difficulties incurred with the use of telemetry by the proximity of other electronic signal transmission equipment, such as radio transmitters or receivers. Improper electrode attachment could also present problems during administration. These difficulties could be circumvented, however, by proper planning and training on the part of the administrator.

6.3.4.2.6 Equipment/Materials Needed for Administration

The equipment and materials needed for application of telemetry techniques include the following:

(1) Telemetry Receiver
(2) Telemetry Transmitter
(3) Telemetry Transmitter Adapter Kit for each physiological function measured
(4) Telemetry Electrodes for each parameter measured

6.3.4.2.7 Personal Effects on Applicants

As with other forms of physiological measurement, telemetry techniques might be viewed by some applicants as constituting an invasion of privacy. This procedure might be seen as less objectionable than biofeedback equipment or the polygraph, however, since the individual is not directly attached to any machines.

6.3.4.2.8 Compliance with Legal Issues and EEOC Guidelines

This review did not reveal any court cases which centered on the use of telemetry procedures. Because this technique has not yet been subject to much evaluative research, it has yet to demonstrate adequate reliability and validity. Due to this lack of empirical data, as well as the technique's potential for being viewed as an invasion of privacy, it must be concluded that compliance of telemetry techniques with legal guidelines remains questionable at this time.

6.3.4.2.9 Confidentiality of Measurement Technique Results

The data obtained through telemetry procedures are personal in nature, and should therefore be maintained in a confidential manner if utilized for selection purposes. Only the administrator and those persons clearly involved in the selection process or personnel matters should have access to this information.

6.3.4.2.10 Susceptibility to Faking

Unlike the previously discussed methods of physiological measurement, telemetry procedures are designed to be utilized in situations where the monitored individual is engaging in some form of motor activity. For this reason, overt body movements would not generally be expected to alter the results, even if this were an applicant's intention. Thus, telemetry appears to be less susceptible to distortion than some of the other methods of physiological measurement.

6.3.4.2.11 Labor Relations Considerations

This review did not reveal any labor disputes which have centered on the use of telemetry. If this procedure were to become the object of a labor dispute at some future date, it is possible that the use of this technique could be considered a violation of privacy by labor organizations.

6.3.4.2.12 Reevaluation Considerations

In accordance with EEOC Guidelines, an individual screened out on the basis of telemetry results should be provided with an opportunity for reevaluation at a later date. This provision would appear to be especially appropriate to applicants who were not feeling well, or who, for some other reason, exhibited abnormally high arousal levels on the initial testing occasion.

6.3.4.2.13 Applicability of Telemetry to Selection of Nuclear Facility Personnel with Respect to Emotional Instability

At the present time, the evidence regarding telemetry techniques is inconclusive; therefore, such methods of physiological measurement cannot be regarded as applicable to nuclear facility selection procedures at this point. Because of the potential value of telemetry in measuring stress responding during simulation situations, further research on the application of these techniques would seem to be a worthwhile direction to pursue.

6.3.5 Voice Stress Analyzers

6.3.5.1 Overview

Voice stress analyzers were originally developed for purposes similar to that of telemetry techniques. Williams and Stevens (Ref. 47) reported the development of equipment designed to detect changes in pilots' vocal utterances during flight, by means of spectrographic analyses of tape-recorded conversations.

Changes in the frequency contour of a pilot's speech pattern were presumed to be correlated with conditions of psychological stress. The currently available voice stress analyzers operate on the same principle, yielding spectrograms which are visually inspected for microtremors in the voice. The microtremors are thought to be attenuated by emotional arousal.

The current major application of the voice stress analyzer is in detection of deception. The technical advantage held by this technique over the use of the polygraph for lie detection is its not necessitating that the individual be attached to any equipment.

6.3.5.2 Relevant Considerations

6.3.5.2.1 Reliability

With respect to voice stress analyzers, traditional evaluation studies have centered on the internal consistency and inter-rater approaches to the measurement of reliability.

Smith (Ref. 48) conducted two studies designed to evaluate the relative effectiveness of the traditional visual inspection method of scoring the voice stress analyzers, and a method using more objective measurement. The first study examined the tapes of thirty-five professional and non-professional broadcasters which were made from radio broadcasts. The second study employed eighteen subjects with phobic anxiety and fifteen non-phobic subjects who were instructed to

count from one to ten while tape recordings were made of their recitations. In analyzing the technique's reliability Smith pooled the data from the two studies and analyzed the tapes from all sixty-eight subjects using a split-half procedure. The results yielded a reliability coefficient of .39.

Horvath (Ref. 49) conducted a study examining the effects of differences in subject motivation on detection of deception. Two trained evaluators were employed to analyze the voice stress analyzer data collected from sixty-four college students. Inter-rater agreement averaged .65 in the detection of deception for this study.

6.3.5.2.2 Validity: Perspective

With regard to the voice stress analyzer, traditional evaluation studies have centered on the construct validation approach to the measurement of validity. As Montgomery (Ref. 50) pointed out, there has been a lack of methodological consistency and scientific rigor in the research studies which have attempted to validate this technique. In this regard, he cited the subjectivity involved on the part of the examiner in his/her judgment of a person's degree of stress following chart analysis, as well as the more fundamental presumption of the relationship of stress level to the truthfulness of the individual's statements.

Kubis (cited in Ref. 50) conducted a study comparing the polygraph, the voice stress analyzer, and the overall impressions of observers using no equipment in a simulated theft situation. The design of the study utilized a total of 176 participants divided into groups of three, with each group consisting of a supposed thief, lookout, and innocent party. The objective of the instrumentation operators and the observers was to correctly discriminate the individuals within each group. The results showed that the polygraph was successful 76% of the time, the voice stress analyzer 32% of the time, and the observers 55% of the time. According to Montgomery (Ref. 50), this study by Kubis remains the most commonly cited and debated investigation of the voice stress analyzer, and implies a less than random chance validity for this instrument (33% representing random chance in a situation involving three choices). Lykken (Ref. 42) summarized the findings from several other studies and also concluded that the validity of the voice stress analyzer did not surpass random chance figures.

An important point was made by both Montgomery (Ref. 50) and Lykken (Ref. 42) regarding the validity studies of the voice stress analyzer. Both of these authors noted that much of the available research, particularly those studies reporting good accuracy for this technique, has been conducted by individuals who, in some way, have vested interests in the voice stress analyzer. These data must be viewed with caution due to their lack of conformity to proper experimental procedures.

6.3.5.2.3 Cost of Development and Administration

There are currently three major manufacturers of voice stress analyzer equipment. The approximate cost of a complete system purchased from one of these companies is approximately $5,000.

6.3.5.2.4 Administrator: Training and Qualifications

At the present time, there are no existing laws governing the credentials of voice stress analyzer operators. Montgomery (Ref. 50) pointed out that this technique's potential for detrimental effects to individuals tested implies the need for governmental regulation which would require standardized training and licensing of operators. Price quotations from one of the manufacturers of the voice stress analyzer revealed that one week's worth of operator training is included in the cost of the instrument. It seems questionable whether a training program of such short duration could realistically provide potential operators with sufficient instruction for proper administration and interpretation.

6.3.5.2.5 Administration of Measurement Technique: Time and Difficulty

There is no fixed amount of time required for administration of the voice stress analyzer. The length of administration is determined by the operator depending on the purpose of testing.

The major difficulty involved with the use of this technique is the necessity for obtaining a high quality tape recording of an individual's speech for analysis. Poor quality tapes would not be able to be converted into accurate spectrograms by the instrument.

6.3.5.2.6 Equipment/Materials Needed for Administration

Smith (Ref. 48) outlined the equipment necessary for administration of the voice stress analyzer to include the following:

(1) A high quality reel-to-reel or cassette audio tape recorder

(2) The voice analysis instrument

(3) Recording paper and ink

6.3.5.2.7 Personal Effects on Applicants

As with other forms of physiological measurement, the voice stress analyzer may be viewed as an invasion of privacy by some individuals, particularly if it is used in questioning a

person about personal background information, as in the case of pre-employment testing. An even more serious issue was discussed by Montgomery (Ref. 50) who noted that there is a significant danger in the possibility that tape recordings of an individual's speech might be obtained without his/her knowledge and then be subjected to voice stress analysis. In this respect, this technique appears unique among the available methods for physiological measurement, as its unobtrusiveness represents a significant departure from the other methods.

6.3.5.2.8 Compliance With Legal Issues and EEOC Guidelines

The major legal issues involving the use of the voice stress analyzer relate to the possibilities of invasion of privacy and Fifth Amendment violations (Ref. 50). The possibility of self-incrimination would be particularly significant with this technique's use in employment selection as a truth verification procedure for background information supplied.

Another issue involving the legalities of the use of the voice stress analyzer for employment screening relates to its potential for discrimination. Montgomery (Ref. 50) noted that this technique, like the polygraphic lie detection test involves the possibility of discrimination by the examiner on the basis of gender or other demographic characteristics of subjects tested. For these reasons, compliance of the voice stress analyzer with legal issues and EEOC guidelines remains questionable at this point.

6.3.5.2.9 Confidentiality of Measurement Technique Results

As previously discussed, there are serious implications involving the use of the voice stress analyzer due to the fact that tape recordings of an individual's voice can be obtained without his/her knowledge. Assuming, however, that a job applicant was informed of the use of this procedure, the results should be treated as confidential in nature, and should only be available to the examiner and those in charge of employment selection and subsequent personnel matters.

6.3.5.2.10 Susceptibility to Faking

Smith (Ref. 48) discussed an earlier study by Smith which showed that hyperventilation had an effect on voice stress analysis patterns and that this effect appeared unrelated to psychological stress. This suggests that, as with polygraphic lie detection tests, individuals may intentionally distort their response patterns by actions designed to alter normal physiological arousal patterns. Thus, job applicants could, if so desired, invalidate the results of this procedure.

6.3.5.2.11 Labor Relations Considerations

This review did not reveal any labor disputes which have centered on the use of the voice stress analyzer. As with the polygraphic lie detection test, however, it is quite possible that labor organizations would consider this technique to be an invasion of privacy.

6.3.5.2.12 Reevaluation Considerations

Compliance with EEOC guidelines dictates that a job applicant screened out on the basis of results of a selection procedure should be permitted to undergo this procedure again. As with other physiological measurement techniques, the voice stress analyzer appears particularly appropriate for reevaluation consideration, due to normal fluctuations in patterns of physiological functioning.

6.3.5.2.13 Applicability of Voice Stress Analyzers to Selection of Nuclear Facility Personnel with Regard to Emotional Instability

At the present time, the voice stress analyzer does not seem to have adequate supportive documentation to justify its inclusion in a selection procedure. Additionally, the questions regarding this technique's legal compliance add further reservations to its consideration.

Since this technique has some practical benefits in that it doesn't require electrodes and cables for administration, it might be advantageous for further research to investigate applied uses of this technique for measuring stress arousal in employment selection procedures. If perfected for such a purpose, this technique would seem to have potential for use in conjunction with simulated stress situations in detecting excessive stress arousal as an index of possible emotional instability.

6.4 Situational Simulations

6.4.1 Overview

Situational simulations are a standardized procedure in which a participant, or participants, are evaluated on their ability to perform work-related tasks under conditions that closely simulate the actual work environment (Ref. 51). Simplistic types of situational simulations include typing tests and driving tests. Situational simulations can be used for employee development, training, or selection.

The first major use of situational simulations was conducted by the American Telephone and Telegraph Company in 1956 and was essentially a research study (Refs. 52 and 53). The purpose of this longitudinal research study, referred to as

the Management Progress Study, was to develop employee's career histories and to compare these histories to future performance (Ref. 54). Career histories were based on data gathered from clinical interviews, projective tests, and situational simulations (referred to as simulation exercises) which included work samples, paper-and-pencil tests, and participation in group problems and group discussions. A team of evaluators, consisting primarily of professionally trained persons, reached consensus regarding the evaluation of each of the employees. The criterion measures were periodic performance evaluations. Correlations between the simulations and on-the-job performance measures have been quite high (Ref. 54).

The first nonresearch use of an assessment program containing simulations was carried out by Michigan Bell Telephone Company in 1958 (Ref. 55).

Michigan Bell's assessment program is the prototype of the simulations used in the Bell System today, as well as the model for those used in other organizations (Ref. 54). The use of the approach has grown rather rapidly over the last twenty-five years.

As the use of situational simulations, also referred to as assessment centers, blossomed, a Task Force on Assessment Center Standards was established. This Task Force set forth the minimal standards of an assessment center, and the interested reader should refer to the appropriate reference (Ref. 56).

The present-day assessment center consists of a standardized evaluation of behavior based on multiple inputs. Multiple trained observers and simulations are used. Judgments about behavior are made, in part, from the specially developed situational simulations. These judgments are pooled by the observers at an evaluation meeting during which all relevant assessment data are reported and discussed and the observers agree on the evaluation of the critical factors relating to the job.

There are some problems associated with the use of situational simulations, including:

- Demand on Organizational Resources: The use of situational simulations may place a heavy toll on internal resources, i.e., personnel. That is, given that the evaluatees and/or evaluators may be from within the organization, the organization will have to tolerate a number of people being away from their regular job duties during the actual evaluation period. This can become somewhat costly, as the average evaluation may take approximately three days.

- Appropriateness of the Actual Situational Simulations: A critical issue is the degree to which the simulations reflect the demands and content of the job as reflected by the job analysis. Simulation development is a time-consuming activity requiring a proper understanding of the job. Without the proper understanding of the job, the simulations will be lacking in terms of content validity.

- Standardization of Procedures: Cohen (Ref. 57) stated that the lack of standardization may invalidate an otherwise useful selection instrument. According to Cohen (Ref. 57) some of the areas of concern include the time permitted each participant to complete each individual simulation and the instructions given to the participants. Cohen (Ref. 57) also stated that the skill of the evaluators should be as uniform as possible, thus ensuring that the evaluation of the various candidates is equally uniform. This therefore necessitates careful training of the evaluators.

In summary, while these concerns are not insurmountable, it is clear that they must be seriously considered if the situational simulation approach is to be used to its fullest potential.

6.4.2 Relevant Considerations

6.4.2.1 Reliability

With respect to situational simulations, traditional evaluation has centered on test-retest reliability, internal consistency reliability, and inter-rater reliability.

The Michigan Bell Telephone Company examined test-retest reliability by evaluating 85 management employees at two different points in time and obtained a correlation coefficient of .73 (Ref. 58).

Other studies of test-retest reliability have reported correlation coefficients ranging from .06 to .74, with the average being .35 (Refs. 52, 59, and 58).

Internal consistency, as it applies to assessment centers, can be defined as consistent performance within an individual skill across all of the situational simulations.

Archambeau (Ref. 60) examined internal consistency reliability and reported coefficients ranging from .36 to .81, with an average of .61. McConnell and Parker (Ref. 59) investigated internal consistency reliability and reported coefficients ranging from .85 to .98, with an average of .88 for six separate assessment centers.

Hinrichs and Haanpera (Ref. 61) reported on the internal consistency reliability of assessment centers conducted in eight countries. They found that the internal consistency coefficients ranged from .04 to .73 with the average being .49. Additional analyses indicated an average internal consistency coefficient of .67. Of particular interest was the skill of resistance to stress, which may be viewed as an index of emotional stability. The internal consistency coefficient for this skill was .66.

Inter-rater reliability is of major concern, with respect to situational simulations, as multiple evaluators are used in making judgments about the evaluatees.

Research studies have reported inter-rater reliabilities in assessment centers ranging from .60 to .98 (Refs. 53, 62, 63, 64, and 65). In a study by Denning and Grant (Ref. 66), the inter-rater reliabilities ranged from .50 to .82, with the inter-rater reliability for the skill "stress tolerance" (somewhat similar to emotional stability) being .73.

In studies which reviewed the research on inter-rater reliability within a single simulation exercise, Bray and Grant (Ref. 53), Carleton (Ref. 67), Finley (Ref. 68), Greenwood and McNamara (Ref. 69), and Thomson (Ref. 65) have reported average coefficients ranging from .26 to .92, with the overall average being .57.

It has been discussed and documented that when substantial time is allocated to assessor (evaluator) training the inter-rater reliability is substantially increased (Refs. 53, 61, 65, and 70).

6.4.2.2 Validity

With respect to situational simulations, traditional evaluation has centered on criterion-related validity, content validity, and construct validity.

In the Management Progress Study, which covered a time span of 8 years, over 400 candidates were evaluated, including both college and non-college employees. Bray and Grant (Ref. 53) reported that the predictive validity correlations between assessment center performance and level of management reached was .44 for the college group and .71 for the non-college group, both statistically significant. Correlations obtained between the assessment center ratings and salary increments for over 200 of the employees ranged from .39 to .52, all of which were statistically significant.

In an AT&T study in which nearly 6,000 individuals were assessed to determine correlations between their assessment center ratings and management progress, the predictive validity coefficient obtained, .44, was statistically significant (Ref. 58).

Other AT&T studies have also reported the success of the assessment center at predicting future job performance (Refs. 71, 72, 73, 74, and 75).

Studies at IBM have also reported strong positive relationships between assessment center performance and performance on the job (Refs. 76 and 77). In a study of 94 lower and middle-level managers at IBM, Wollowick & McNamara (Ref. 78) reported that the correlations between assessment center results and job position level three years later ranged from .29 to .63, with seventy-five percent of these correlations being significant.

In four separate studies conducted at Standard Oil of Ohio, criterion-related validity coefficients ranged from .63 to .65 for the correlations between assessment center results and supervisors's ratings of on-the-job performance (Refs. 68 and 65). Strong relationships between assessment center performance and on-the-job performance have also been established at General Electric (Ref. 79), Sears Roebuck (Ref. 80), and Wickes Corporation (Refs. 81 and 82).

In a review of 23 previous studies on the assessment center process, Byham (Ref. 83) reported that the correlations between assessment center predictions and various on-the-job performance measures ranged as high as .64. In a review of eighteen research studies on the assessment center approach, it was reported that the average criterion-related validity was .40 when number of promotions was used as the criterion measure and .63 when supervisor's rating of the participant's promotion potential was used (Ref. 84) as the criterion measure.

With respect to emotional instability, there have been some studies which have attempted to relate skills assessed such as "resistance to stress" or "tolerance to stress" to future measures of on-the-job criteria. Despite generally supportive results, not all of these studies have used future performance on the job as criterion measures. One study conducted by IBM* utilized the predictive criterion-related model with approximately one hundred professional financial personnel. One of the skills measured in the simulations was "resistance to stress." The criterion measure was relative degree of advancement or promotions after a thirteen-year

* Personal communication with H. Schwartz, December 17, 1980.

period. This study demonstrated statistically significant results in support of situational simulations (including stress tolerance skills) as being valid predictors of on-the-job performance.

With respect to content validity, it can be determined by comparing the results of a carefully conducted job analysis and the content of the situational simulations. That is, the content validity of a measurement instrument is established by comparing the instruments to the findings of a relevant job analysis. With proper care, situational simulations can be supported through a content validity basis.

Construct validity can be supported on the basis of an analysis of convergent validity and discriminant validity, and this method has been relied upon with respect to situational simulations.

In general, the research reported indicates that situational simulations do show convergent validity, but do not show strong evidence of discriminant validity (Refs. 58, 59, 85, 86, and 65); thus, overall, there is mixed support with respect to the construct validity of this approach.

6.4.2.3 Cost of Development and Administration

In studies of assessment center costs, costs have ranged from $400 to $1000 per evaluatee (Refs. 87, 88, and 89). Recommendations have been made for reducing costs including the consortium approach (Refs. 90 and 91) where several organizations in conjunction with one another support an assessment center program.

6.4.2.4 Administrator Training and Qualifications

Of critical concern in assessment center applications are the qualifications and training of assessors/evaluators. For this reason, this subsection will be devoted to assessor related concerns as opposed to focusing on administrator issues. The question becomes, do evaluators need to be professional psychologists or can non-psychologists be properly trained to be evaluators?

Investigations of inter-rater reliabilities and other statistical indices have shown that evaluators need not be professional psychologists (Ref. 69), but that they do need to be properly trained (Ref. 70). This means that the training of evaluators becomes critical since inadequate assessor training can result in invalid and unreliable evaluations (Ref. 92).

The reason that evaluator training is so essential is that it is critical that evaluators apply the same set of standards to all participants (Ref. 93). In order to ensure fairness

in testing, all participants must be evaluated against the same set of criteria. In surveying various organizations, evaluator training may last anywhere from six hours to three weeks (Ref. 51).

An explanation for this large variation in the amount of time devoted to evaluator training is that the major concern is the quality, not the quantity, of training. In this respect, Frank and Whipple (Ref. 93) established an evaluator certification program at General Telephone of Florida. In this program, evaluators, after receiving training, are themselves evaluated on their performance in situational exercises that closely simulate the critical tasks performed by 'real' evaluators.

The Task Force on Assessment Center Standards directly addressed this issue when they established the Standards and Ethical Considerations of Assessment Centers (Ref. 56). The considerations state that multiple assessors must be used and that an individual must receive proper training prior to serving as an assessor.

6.4.2.5 Administration of Situational Simulations: Time and Difficulty

The time needed to administer situational simulations generally varies from one half day to five days, depending on the number of simulations, participants, and evaluators that are involved.

6.4.2.6 Equipment/Materials Needed for Administration

Depending on the type of simulation being used, the equipment needed may range from stationery to sophisticated special machinery. With respect to the nuclear facility programs, the complexity and cost of the equipment needed would depend on whether or not the nuclear facility decided to build simulations of the special equipment used by personnel (e.g., central security alarm system).

Even if the nuclear facilities chose not to simulate the special equipment used on the job, specially designed materials would still be needed. Since situational simulations are designed to actually simulate the position of the focal personnel, the scenarios must be carefully developed, based upon an accurate and thorough job analysis of the focal position.

6.4.2.7 Personal Effects on Applicants

Researchers have found that participants' reactions to the situational simulation process has been positive both immediately following participation and after at least a one-year time period (Refs. 94, 95, and 96). Nirtaut (Ref. 95) surveyed 47 organizations using situational simulations and

found that none of the organizations reported negative participant reaction and that this was true regardless of the level of employee assessed, the use made of the information generated by the assessment center, or the weight given to the information generated by the assessment center in the selection decision.

Slivinski et al. (Ref. 96) and Quaintance (Ref. 97), found similar results. Of particular interest was the Quaintance study (Ref. 97) where she investigated the reactions of unsuccessful participants to the assessment center process, and found that the participants' reactions to the assessment center and the feedback were very positive.

6.4.2.8 Compliance with Legal Issues and EEOC Guidelines

The assessment center process has been challenged and upheld in the courts, (Refs. 91, 98, 99, 100, and 101). The use of the assessment center was specifically sanctioned in a consent decree involving AT&T and the EEOC (Ref. 102). Similarly, in a case involving the City of St. Louis and the Firefighters Institute, the court issued a decree stating that a technique which uses individual and group exercises to simulate job responsibilities, with assessors rating the candidates' performance (i.e., an assessment center) can be used in making promotional decisions (Ref. 103).

Empirical research has shown that properly constructed situational simulations are equally valid for predicting performance regardless of the individuals' race or sex (Refs. 81, 73, 74, 104, and 105).

6.4.2.9 Confidentiality of Situational Simulation Results

As with all other measurement instruments, the confidentiality of the participants' performance results is critical (Ref. 51). Since the assessment center process involves multiple evaluators, it may be more difficult to control the confidentiality of results for the simulation exercises than it would be for other measurement instruments.

6.4.2.10 Susceptibility to Faking

The question of whether performance in the situational simulations can be faked has been empirically investigated. Specifically, researchers have examined whether prior knowledge of the assessment center process has any effect on the participants' subsequent performance. In this respect, the empirical investigations of faking have actually concerned themselves with examining whether coaching (i.e., transmitting of pertinent information) has any effect on subsequent assessment center performance. Studies generally have shown no substantial practical effects of coaching (Refs. 66, 106, 107, and 108).

The most direct study on the effects of coaching, or foreknowledge of the assessment center process, was conducted by Struth and Frank (Ref. 109). The purpose of this study was to investigate whether performance in situational simulations was enhanced for participants who had previously participated in evaluator training sessions and served as evaluators in an assessment center. The results of this study indicated that training and serving as an evaluator did not significantly affect subsequent assessment center performance both with respect to individual skill performance and overall performance. Given the extent to which assessor training exposes individuals to the intricacies and subtleties of the assessment center process, this study certainly speaks to the assessment center's resistance to coaching effects.

6.4.2.11 Labor Relations

The issue of union and employee acceptance of situational simulation in selection is related to the legal and EEOC concerns of the simulations and their effect on the participants. The job-relatedness of the simulations has helped to make them acceptable to participants and employees (Ref. 51). Therefore, it does not seem that the use of situational simulations will cause any severe labor relation problems, and they have been used successfully within union settings.

6.4.2.12 Reevaluation Considerations

A key consideration here is that the skills measured by the simulations may be learned through normal work experience, actual real-life experiences, and/or self-teaching. Thus, it would be appropriate to allow individuals to be reevaluated after a specified time period has elapsed.

6.4.2.13 Applicability of Situational Simulations to Selection of Nuclear Facility Personnel with Regard to Emotional Instability

Originally, situational simulations were used for the purpose of evaluating, selecting, and training technical personnel (e.g., typists, pilots). Since many nuclear facility personnel are oftentimes required to monitor machines, such as the central security alarm system, the use of simulations would be appropriate for the selection and training of such personnel. In this context, consideration should probably be given to the cost involved in simulating machinery such as the central security alarm system panel. A decision would involve calculating whether the savings resulting from the selection and enhancement of training would outweigh the cost.

As simulation became more sophisticated, the use of in-baskets and business games was introduced to simulate non-technical positions and the skills required of personnel in these

positions. These and other types of simulations have been used in the selection and training of managerial and sales personnel, as well as police and fire department managerial personnel. Other applications of situational simulations, for instance with patrolman, have used other types of simulations, such as intervening in hypothetical domestic confrontations (Ref. 110). Many simulations have included evaluations of the participants' emotional stability or resistance to stress. This has considerable promise in terms of nuclear facility personnel selection.

In conclusion, it seems that both technical and non-technical simulations may be appropriate for the evaluation of nuclear facility personnel.

6.5 Weighted Application Blanks

6.5.1 Overview

The purpose of an application blank is to obtain information about applicants which may be used for selection purposes. Generally, this information is biographical in nature. Examples of the types of information collected from an application blank include: work history, educational history, military record, and marital status. Application blank information or biographical data are often referred to as biodata. Usually the biodata items call for factual data, but items which tap attitudes, feelings, or value judgments are sometimes included (Ref. 111).

Basically, there are two types of application blanks, the traditional application blank and the weighted application blank (WAB). One difference between the two types is that a weighted application blank, also referred to as a scored autobiographical data blank, consists of standardized and multiple choice questions while this is not necessarily the case with traditional forms. The standardized questionnaire allows for objective comparisons between applicants.

A second, and more major difference between traditional and weighted application blanks concerns how the information is used by the organization. That is, whereas the information received on the traditional application blank is often used in making selection decisions, it is generally used in a non-empirical way. In contrast, questions used on the weighted application blank are based on the results of criterion-related validity studies. From the validation studies, the relative predictability of the items can be determined. That is, the items with high validity coefficients are better predictors than those with low coefficients. With the criterion-related validity coeffficients obtained, the items are weighted according to their predictability; items which are high predictors are weighted stronger than those which are moderate or low predictors.

6.5.2 Relevant Considerations

6.5.2.1 Reliability

With respect to weighted application blanks, traditional evaluation has centered on test-retest reliability. To a lesser extent the internal consistency reliability of weighted application blanks has also been evaluated.

Test-retest reliabilities have ranged between .82 and .90 (Refs. 111, 112, and 113). It must be noted that these test-retest studies all used students as subjects. Care should be taken when relating these results to employee applicants.

Although the internal consistency method has not been a primary way of measuring the reliability of weighted application blanks, Baehr and Williams (Ref. 114) computed Kuder-Richardson 20 reliabilities in a study using 680 vocationally different males. The authors reported reliability coefficients ranging from .43 to .76.

6.5.2.2 Validity

With respect to weighted application blanks, traditional evaluation has centered on criterion-related validity.

Results of the use of weighted application blanks by the U.S. Military during World War II included reports on criterion-related validity. The validity coefficients for predicting the success in training of Air Force student pilots averaged from .35 to .40, while those for predicting navigator success ranged from .25 to .30 (Ref. 115). The U.S. Army reported a validity coefficient of .45 for predicting leadership in personnel attending Officer Candidate School (Ref. 116). In predicting the leadership qualities of 2003 members of the Reserve Officer Training Corps, Roy, Brueckel and Drucker (Ref. 117) reported a validity coefficient of .26.

Standard Oil Company of New Jersey conducted an in-depth study of the criterion-related validity of scored biographical data in predicting overall job performance. The validity coefficients obtained were .45, .30, .27, .40, and .43 for predicting the performance of skilled craftsmen, engineers, technical personnel, office personnel, and supervisory personnel respectively (Ref. 118).

Richardson, Bellows, Henry, and Company[1] (see (Ref. 111) reported the results of a study involving 477 black, disadvantaged males and 225 black, disadvantaged females, from three different cities, as subjects. The criteria were tenure on the job after three months and after six months. The validity coefficients obtained ranged from .37 to .53.

Buel (Ref. 119) conducted a study investigating the validity of weighted application blanks for predicting voluntary turnover among female clerical employees. This study used a weighted application blank consisting of 16 items and a sample of 72 employees. The reported criterion-related validity was .49.

Other criterion-related validity studies have reported that the weighted application blank validly predicts job tenure (Refs. 120 and 121) and job performance[2] (Refs. 120, 122, and 123).

6.5.2.3 Cost of Development and Administration

Glueck (Ref. 124) stated that it has been estimated that it takes about 100 hours to develop a weighted application blank. Once the blank has been developed, the on-going cost of using it is quite minimal.

6.5.2.4 Administrator: Training and Qualifications

Once the weighted application blank is developed, only a minimal amount of time and effort is needed to train a person to administer and score the instrument.

Weighted application blanks can either be scored by machine or, with a minimum expenditure of time, effort, and money, employees can be trained to administer them.

---

1 Richardson, Bellows, Henry, and Co., "Predicting Job Tenure Among ES Applicants and Program Tenure Among WIN Clients Through the Use of Biographical Information, Washington, D.C. (1971).

2 B.S. Gantz, C. Erickson, and R.W. Stephenson, "Test Prediction of Promotion Rate and Job Satisfaction for Scientists and Engineers," Paper read at California State Psychological Association, San Diego (1969).

6.5.2.5 Administration of Weighted Application Blanks: Time and Difficulty

The administration of a weighted application blank is virtually identical to the administration of an non-weighted (traditional) application blank. The administration time depends on the total number of items on the application blanks. The average administration time will typically range from fifteen to thirty minutes.

6.5.2.6 Equipment/Materials Needed for Administration

No unique or extraordinary equipment or materials are needed for the administration of weighted application blanks.

6.5.2.7 Personal Effects on Applicants

In general, job applicants will not react adversely to being required to complete an application blank, as long as they feel that the questions being asked do not invade their privacy and are job-related. That is, although applicants will not resent being asked questions regarding their military records, they may feel offended by the question concerning personal habits such as use of alcoholic beverages.

Since virtually all organizations require applicants to complete some type of application form, applicants have come to expect that they will be asked to complete a form. The widespread use of application blanks has contributed to their overall acceptance by job applicants.

6.5.2.8 Compliance with Legal Issues and EEOC Guidelines

To the extent that an application blank weights items concerning race, sex, religion, or national origin, such that the total score derived adversely affects the members of any protected class, decisions based on the application blank are potentially illegal. Consideration must be given to items which specifically ask about, as well as items which are highly correlated with, an applicant's race, sex, religion, or national origin.

Pace and Schoenfeldt (Ref. 125) suggested that the decision of what items to include in the weighted application blank be based on the results of a thorough and accurate job analysis (i.e., content validity) and that these items then be correlated with a criterion measure (i.e., predictive validity). Such an approach should produce an instrument that is both valid and defensible in terms of the provisions of the equal employment opportunity guidelines.

6.5.2.9 Confidentiality of Measurement Technique Results

The confidentiality of the results of the weighted application process depends upon the policies of the particular organization. It is appropriate to limit access to these results to as few people as possible. Information contained on the application blank, as well as the resultant weightings, may contain data that are personal in nature.

6.5.2.10 Susceptibility to Faking

Faking can be a problem with weighted application blanks. Job applicants can falsify information reported on their applications. Therefore, the weighted application blank must be constructed in such a way that items are verifiable. Complete verification checks should be made of all of the information contained on the weighted application blank, as well as for information obtained from a traditional application blank.

Informing applicants that the information on the application blank will be verified may lead applicants to respond more honestly. Applicants may be reluctant to falsify information because of the fear of being detected.

6.5.2.11 Labor Relations Considerations

Employees, or unions, may object to the use of weighted application blanks if they feel that the biographical information includes some non-job related data. Thus, careful control of the information obtained is a mandatory requirement in terms of privacy considerations.

6.5.2.12 Reevaluation Considerations

Since some of the information reported on the weighted application blank is susceptible to change over time (e.g., marital status, address, age), it would be appropriate to allow applicants who were previously denied employment based on their application blanks, to complete a new application blank after an acceptable amount of time (approximately one year).

Also if the weightings for the individual items on the application blank are changed, or the criterion is changed, applicants who were previously denied employment based on their original application blank responses should be given a second opportunity.

6.5.2.13 Applicability of Weighted Application Blanks to the Selection of Nuclear Facility Personnel with Regard to Emotional Instability.

The overall value of weighted application blanks for predicting emotional instability may be questionable given potential

difficulty in documenting correlations between item responses and actual indices of emotional instability on the job. That is, given the non-availability of criterion measures, it may be difficult to develop proper weightings.

6.6 Clinical Interviews

6.6.1 Overview

Despite several variations in approach, the clinical interview is basically a method of personality assessment consisting of the exchange of verbal information between an individual and a trained psychiatrist or clinical psychologist. The clinical interview is the most long-standing and widely used method of psychological assessment. Its origin as a formal technique can be traced to the process of free association, developed by Sigmund Freud (cited in Ref. 126). In using this approach, Freud conducted individual sessions with his patients, in which he encouraged them to say whatever came to mind, without regard to social standard or self-censoring.

The clinical interview has been subjected to a great deal of further study and specification since Freud's time. Rotter and Hochreich (Ref. 2) outlined the three basic approaches to clinical interviewing. The first, free interviewing, is characterized by a situation in which the interviewer says as little as possible, asking only open-ended questions and encouraging the person to discuss whatever is on his mind. Although this format tends to be the least-threatening kind, it is often quite time consuming and may provide little pertinent data.

The second approach is the directed interview, in which the interviewer attempts to obtain specific kinds of information pertaining to presenting problems (i.e., those problem behaviors or feelings which prompted the individual to request or be recommended for a clinical interview). In this case, the interviewer asks direct questions designed to gather information regarding these problems. This method typically provides more information than does the interviewing approach, but may be more intimidating to the individual and thereby restricts his/her spontaneity.

The third approach is the structured interview, in which a standard procedure is followed, with the interviewer asking the same set of questions in a specified order, of all the interviewees. These questions are designed to probe the major aspects of an individual's personal and inter-personal functioning such as self-concept, and family relationships, in an effort to detect problem areas. Although this technique is likely to place even more restrictions on an individual's volunteering of information, it has the advantages of being less time consuming, providing more uniform interviewing conditions, and establishing a basis for comparison among individual's interviewed.

A more recent development in the area of clinical interviewing is the behavioral interview. This approach was developed in an attempt to more systematically gather quantitative and situation-specific information pertaining to the environmental events associated with an individual's behaviors, as well as the person's thoughts and beliefs regarding these events (Ref. 127). Another objective of this technique is to provide data sufficient for adequate research investigations relating to clinical interviews, which have, historically, been quite minimal.

If a clinical interview is to be used as part of a routine personnel selection system, a standard interview format must be used in order to provide for a uniform screening procedure for all applicants.

6.6.2 Relevant Considerations

6.6.2.1 Reliability

With respect to clinical interviews, traditional evaluation studies have centered on the inter-rater approach to measuring the reliability of clinical interviews.

The research utilizing the inter-rater approach to measure the reliability of clinical interviews has utilized two types of methodologies. The first type has involved comparisons of the original interviewer's ratings of an individual's emotional status with those ratings provided by clinicians who have subsequently viewed videotapes of the original interview. The second approach has compared the ratings from multiple interviews with one individual involving a series of different interviewers.

The research cited in these studies (Abrams and Rheed, cited in Ref. 128; Ref. 129; and Adams, cited in Ref. 128) has not demonstrated good support for the inter-rater reliability of clinical interviews; however this is probably due in part to methodological problems involved. The format of the structured clinical interview lends itself to the inter-rater reliability approach and well-designed studies should be conducted for this purpose.

6.6.2.2 Validity

With regard to clinical interviews, traditional evaluation studies have centered on the criterion-related approach to measuring validity.

In regard to clinical interviews, Eisenberg, Kent, and Wall (cited in Ref. 128) noted that it is quite difficult to measure statistical validity. This is true because much of the available literature is distorted by artifacts (e.g., the interviewer's mood on the day of the interview), the tendency

toward the use of only concurrent criteria, and conclusions which are not representative of the objective data obtained. Crosby (Ref. 128) suggested that the lack of statistically significant validity data reflects the confounding effects of several non-uniform variables. These include interviewers' rating being biased by their sometimes having access to biographical information and/or psychological test results prior to the interview, as well as factors associated with the interviewees, such as whether or not they have ever had a previous psychiatric interview. Thus, this review revealed no statistically significant data resulting from studies which attempted to compare interviewers' ratings with on-the-job criteria.

Although the construct validity approach would be a viable method of investigating the validity of clinical interviews, this review did not reveal any major studies. The lack of major studies may be due to the fact that the most obvious measurement technique to use in obtaining construct validity data is personality tests. In practice, however, clinicians almost always combine the data from personality tests and clinical interviews to arrive at their conclusions. Thus, comparisons of the data resulting from these techniques have undergone relatively little investigation, leaving this a fruitful area for future research.

6.6.2.3 Cost of Development and Administration

A clinical psychological or psychiatric interview generally ranges from $50.00 to $100.00 per interview if a private practitioner's services are utilized. If the interviewer is permanently on the staff of the organization that desires such services on a regular basis, the cost factor would presumably be less.

6.6.2.4 Administrator: Training and Qualifications

Administration and interpretation of a clinical evaluation should be conducted by a licensed psychologist or psychiatrist to ensure the interview is conducted in a prescribed professional manner, and that ethical and legal guidelines are met.

6.6.2.5 Administration of Measurement Technique: Time and Difficulty

The time required for a clinical interview generally ranges from forty-five minutes to an hour and a half. There are no obvious inherent difficulties associated with this technique.

6.6.2.6 Equipment/Materials Needed for Administration

No equipment is normally necessary for administration of a clinical interview unless a tape-recording is made (which would require written consent of the interviewee).

6.6.2.7 Personal Effect Applicants

Due to the apprehension frequently associated with being asked to discuss "personal" matters with a mental health professional, there exists a possibility of some adverse reactions to the interview (e.g., fear, anxiety, anger, etc.) in a job applicant who is required to undergo this procedure. However, a skilled professional should be adept at first establishing good rapport with interviewees, thus making them feel more comfortable with the situation.

6.6.2.8 Compliance with Legal Issues and EEOC Guidelines

If it can be demonstrated that the interview measures of emotional instability are related to job performance in a nuclear power facility, the inclusion of a clinical interview in a selection procedure would not appear to violate EEOC guidelines. However, if certain directed questions during the conducting of the interview cannot be demonstrated to relate to the construct of emotional instability and to the job, both EEOC Guidelines and Right to Privacy Regulations might be violated.

6.6.2.9 Confidentiality of Clinical Interviews

The American Psychological Association's Code of Ethical Standards (Ref. 7) restricts the transmission of diagnostic results of clinical evaluation to another qualified professional. These results may be interpreted in lay terms for transmission to other individuals with the applicant's written consent, specifying the recipient(s) of the information.

6.6.2.10 Susceptibiity to Faking

A job applicant is likely to attempt to make a good impression during a clinical interview, particularly if he/she is aware that the results will contribute to the selection decision. Therefore, the potential for untruthful statements is highly possible. A well-trained professional should be alert to non-verbal indices suggestive of deception, and should incorporate these data into the Observations Section of his/her Clinical Report.

6.6.2.11 Labor Relations Considerations

The requirement of a clinical interview as part of a selection procedure has the potential to be considered an invasion of privacy, a factor which could result in labor relations conflict. Thus, this speaks to the need to carefully control the data obtained through the interview.

6.6.2.12 Reevaluation Considerations

In accordance with EEOC Guidelines, applicants rejected on the basis of the clinical interview must be provided with the opportunity for reevaluation. Subsequent interviews would be particularly appropriate for individuals who have undergone a course of therapy following feedback of initial interview results. These subsequent interviews should be conducted by a clinician other than the individual who conducted the initial interview with the applicant.

6.6.2.13 Applicability of Clinical Interviews to Selection of Nuclear Facility Personnel with Regard to Emotional Instability

Despite the relatively insufficient statistical support for the clinical interview, it appears to be a vital element in integrating the data derived from other instruments in a selection procedure designed to screen out emotionally unstable job applicants. This technique, if properly structured and conducted by well-trained professionals, appears to be the only means of clarifying inconsistencies in the results of other instruments, as well as probing for further information regarding problems revealed by previously administered components of the selection procedure. It can also assist in determining whether faking revealed by other instruments was due to misunderstanding rather than intentional deception.

The structured clinical interview format offers promise for well-designed reliability and validity studies, which would further substantiate its use in a nuclear facility personnel selection procedure.

6.7 Selection Interviews:

6.7.1 Overview

A selection interview is face-to-face communication used to collect information regarding a job applicant. It differs from both conversation and interrogation because it has explicit purpose and free intercourse between the participants. (Ref. 130).

Whereas the major purpose of a selection interview is the gathering of information about the application, the interview can also serve as a way for the organization to get acquainted with the applicant and a way to present the organization to the applicant. Glueck (Ref. 124) identified three basic types of interview techniques: structured, unstructured, and stress.

In the unstructured, in-depth, or free association method of interviewing, there is only a general topic to be discussed, and the data gathered are a function of the dynamics of the

interaction that takes place. The purpose is to get the applicant to do most of the talking. The applicant has the initiative in how the interview proceeds.

During the structured interview, the interviewer follows, item by item, a predetermined format. The purpose of the interview is to gather and record data. The interviewer controls and directs the interview.

Some interviews are best characterized as a combination of structured and unstructured approach, that is, a hybrid approach. In this type of interview certain data are collected in a structured way (i.e., questions are asked in a specific format) and other data are collected in an unstructured manner (i.e., non-planned, questions used to gain insight into certain areas).

Stress interviews were originally developed during World War II to select potential espionage agents and to assess the ability of the applicants to accept and cope with stress. Glueck (Ref. 124) suggests that it may be appropriate to use stress interviews when there is a high stress content in the job, but that their use even in this situation is questionable. Stress induced during the interview may not directly resemble the stress which occurs on the job, thus an applicant's ability to handle stress in the interview may not necessarily correlate with his/her ability to handle the stress on the job.

Stress interviews can be conducted in either a structured or unstructured manner. The interviewer may always ask the same questions, in the same order, and may always reply to the applicant's questions and responses in the same manner. Or, the interviewer may adjust to the individual situation. The interviewer can introduce new and different stressors. This will allow the interviewer to determine what the applicant's stress tolerance level may be.

6.7.2 Relevant Considerations

6.7.2.1 Reliability

With respect to selection interviews, traditional evaluation has centered on inter-rater reliability.

In a comprehensive review of the literature on the inter-rater reliability of selection interviews, Wagner (Ref. 131) reported reliability coefficients ranging from .23 to .97 (median = .57) for the ratings of specific traits, and from -.20 to .85 (median = .53) for the ratings of overall ability. In total, 96 different traits were rated in the studies reviewed by Wagner, but reliability coefficients were only reported for 29 of those traits. Examples of the traits were: alertness, self-confidence, sociability, social adjustment, and tact.

In other reviews of selection interview research, Mayfield (Ref. 132), Ulrich and Trumbo (Ref. 133), and Webster (Ref. 134), also reported low inter-rater reliabilities. Ulrich and Trumbo (Ref. 133), whose review spanned the research from 1949 to 1964, found only 11 studies that reported data on the reliability of interviewer ratings. Ulrich and Trumbo (Ref. 133) concluded from their review that, "It is apparent, first, that few studies have reported reliabilities and second, that those reported, with few exceptions, are lower than usually accepted for devices used for individual prediction" (p. 108).

More recent research has shown large variation in inter-rater reliability. Palacios, Newberry, and Bootzin (Ref. 135) reported reliabilities ranging from .75 to .91. Minor (Ref. 136) and Stinger and Tyson (Ref. 137) report coefficients ranging from the teens to the .80's. Landy (Ref. 138) reported reliability coefficients consistently in the .80's.

A Key factor in inter-rater reliability is the degree to which there exists common definitions of the traits to be measured. A lack of this common definition may contribute toward some of the low reliabilities that have been reported (Ref. 138).

In this same respect, Hakel (Ref. 139) stated that two other contributors to the low reliabilities are: the differences in the first impressions of various applicants and the differences in the type of data that are first presented in the interview. The basis for Hakel's statement is that a major determinant of an interviewer's final judgment is the interviewer's first impression of the applicant (Ref. 140).

Based more on intuitive than empirical reasoning, researchers and reviewers alike have suggested the use of structured interviews, claiming that inter-rater reliability is generally higher for structured than unstructured interviews (Refs. 124, 141, 133, and 131).

Schwab and Heneman (Ref. 142) found that inter-rater reliability for the structured interview was .79, for the unstructured interview .36, and for the hybrid it was .43.

Inter-rater reliabilities tend to be higher for the structured interview because of its standardization. Also, with a more structured, standardized interview, interviewers are more likely to base their judgments on all of the available information, and less likely to make judgments based on first impressions.

6.7.2.2 Validity

With respect to the validity of selection interviews, traditional evaluation has centered on criterion-related validity, construct validity, and content validity.

Certainly, the typically low reliability of the interview hampers its validity. In order for any measurement instrument to be valid it must first be reliable. In the preceding section it was indicated that a high level of inter-rater reliability for selection interviews has not been established. Therefore, the validity of selection interviews will remain low until the reliability can be improved (Ref. 140).

Carlson* conducted a review of 24 research studies which were concerned with the criterion-related validity of the selection interview. In 9 of these studies the interview was the interviewer's only apparent source of information. Validity coefficients for these studies ranged from .15 to .25.

Landy (Ref. 138) conducted an in-depth study of the criterion-related validity of the selection interview in predicting police performance and concluded, in contrast to Carlson, that the selection interview can validly predict onthe-job performance. However, this study, unlike many other selection interview studies, used multiple interviewers instead of just one interviewer.

Landy (Ref. 138) used both multiple predictor and criterion measures. The predictors were general overall suitability and nine specific interview dimensions; e.g., appearance, experience, communications. The criteria were eight specific performance measures; e.g., judgment, initiative, relations with others. Landy found that the specific interview dimensions were better predictors of the performance dimensions than were the interviewers' opinions of the applicants' general overall suitability. Additionally, it was found that certain performance dimensions (e.g., communication, demeanor, job knowledge) were more accurately predicted than other performance dimensions (e.g., relations with others, initiative, dependability, attitude). These latter dimensions may not be as predictable because of the difficulty involved in behaviorally defining and identifying them.

* R.E. Carlson, "The Current Status of Judgmental Techniques in Industry," Paper presented at the symposium, Alternatives to paper and pencil personnel testing. University of Pittsburgh, May (1972).

The results of Landy's study, specifically that certain traits were more predictable than others, may have particular significance to the selection of nuclear facility personnel with respect to emotional instability. Emotional instability may be considered relatively difficult to behaviorally identify. In commenting on the problem assocated with predicting certain performance dimensions (such as emotional stability), Landy and Trumbo (Ref. 141) stated that prediction may be improved if interviewers are properly trained to recognize behavioral patterns that are indicative of the performance to be predicted, e.g., emotional instability.

One method of ensuring that the selection interview is content valid is to determine the relevant traits and questions through a careful job analysis. However, as Landy and Trumbo (Ref. 141) stated, although in most cases the traits have been chosen on assumptions relating to the job analysis, these assumptions themselves are seldom tested and may be questionable. The traits to be measured may actually be related to the job, but the interviewer may have an inaccurate perception of what type of person possesses these traits or how these traits surface during an interview. In this respect, Landy and Trumbo state that interviewers should have a valid stereotype of a worker who would perform well on the job and that this stereotype should be based on extensive job and worker analyses. In fact, the authors suggested that it may be advantageous to have the interviewers interview successful workers on the job prior to interviewing applicants for the job.

The study by Landy (Ref. 138) mentioned earlier, investigated the construct validity of the selection interview. To do this, he examined the convergent validity of the interview.

The results of the convergent validity investigation were that all the inter-correlations between raters rating the same traits were positive and above .50. From this it was concluded that the interviewers do show a fair amount of agreement in rating the specific traits, (e.g., appearance, communications, education, social sensitivity, sincerity), but there is still room for improvement.

The question of whether the interviewers can discriminate one trait from another is answered by examining the discriminant validity. Higher correlations between different interviewers rating the same traits compared to the same interviewers rating different traits is evidence of the existence of discriminant validity. The results of Landy's study were mixed with respect to discriminant validity (Ref. 141). (People more experienced in interviewing and identifying behavioral traits will be better able to differentiate between the different traits than those people who have a limited amount of experience).

In general, the selection interview has moderate convergent validity, meaning that the evaluations of the different interviewers tend to be similar. However, the discriminant validity of the selection interview is questionable and will vary depending on the experience of the interviewer.

6.7.2.3 Cost of Development and Administration

Glueck (Ref. 124) stated that the cost of the selection interview is a function of the length of the interview times the cost per hour of the interview. The major element with respect to cost per hour is the interviewer's hourly salary. Other costs would include transportation costs and the cost of training the interviewer. The cost of training an interviewer to conduct a structured interview may be more expensive than training for an unstructured interview since the structured interview includes more specific tasks than the unstructured interview. Training an interviewer to conduct a stress interview may be even more expensive. The interviewer, in this case, in addition to conducting a normal interview, must be trained in the method of inducing stress in the interview.

6.7.2.4 Administrator: Training and Qualifications

Interviewers, if they are expected to accurately identify applicant's traits, should undergo training in the techniques of interviewing for the purpose of selection (Refs. 140 and 130). In the case of attempting to predict emotional instability, interviewers should be trained to identify applicants' traits that are predictive of emotional instability.

6.7.2.5 Administration of Selection Interviews: Time and Difficulty

The time required to administer a selection interview varies, but generally lasts from 30 to 60 minutes. In terms of difficulty, once the interviewer(s) are trained, the administration of the interview becomes a relatively simple procedure.

6.7.2.6 Equipment/Material Needed for Administration

There are no special or unique materials needed for the administration of selection interviews.

6.7.2.7 Personal Effects on Applicants

Since almost all organizations utilize selection interviews (Refs. 141 and 143), job applicants are typically accustomed to participating in an interview prior to employment. Additionally, since the interview presents the applicant with an opportunity to meet prospective employers and to view the

actual work place, it may be assumed that the effects of the selection interview on the applicants are, for the most part, positive. This, however, will most likely not be true for stress interviews. That is, applicants will more likely react unfavorably to participation in a stress interview.

6.7.2.8 Compliance with Legal Issues and EEOC Guidelines

The questions asked and the traits measured during the interview should not detract from the reliability and validity of the entire selection process. Establishing the reliability and validity of the selection interview will help ensure compliance with EEOC guidelines.

The questions asked during the interview should be related to on-the-job demands. Asking questions that are not related to job demands may be in violation of EEOC guidelines. Content validity can be used to establish the job-relatedness of the questions asked. Questions asked should not constitute an invasion of the applicants' rights. For this reason, the organization must ensure that the questions asked are job-related and that they are not an invasion of the applicants' privacy.

6.7.2.9 Confidentiality of Selection Interview Results

The problems associated with the confidentiality of selection interview results are related to how well the formal and informal networks of communication are controlled. That is, the more people who have access to the formal results of the selection interview, the more difficult it is for the organization to control the dissemination of an applicant's interview data. The results of the interview should be treated in the same fashion as any testing device.

6.2.7.10 Susceptibility to Faking

Applicants can present either accurate or inaccurate information during the interview. This problem can be reduced by verifying the information obtained during the interview, by telling the interviewees that the information will be verified, and by training interviewers how to identify applicants who may be faking through observation of non-verbal behavior and voice inflections.

6.7.2.11 Labor Relations Considerations

Since, as stated earlier, selection interviews are used by virtually all organizations, they have become an acceptable pre-employement measurement technique. Thus, unless the selection interview violated legal or EEOC guidelines there will most likely be no significant problems with regard to labor relations issues. However, stress interviews may not

be readily accepted by union and/or employees. This possibility must be taken into account when considering the use of stress interviews.

6.7.2.12 Reevaluation Considerations

If an applicant is refused employment based on the selection interview, it would be appropriate to allow the applicant to be re-interviewed after a designated amount of time has elapsed. The reason for this is that the applicant may undergo changes (e.g., additional, relevant work experience; training in skills needed on the job) that may affect his/her suitability for the position.

6.2.7.13 Applicability of Selection Interviews to the Selection of Nuclear Facility Personnel with Regard to Emotional Instability

One immediate consideration is whether the stress interview could be effectively used to measure the emotional stability of applicants, as a certain degree of stress would be present on the job. In evaluating this consideration, nuclear facilities should concern themselves with the following questions:

(1) Does the demonstration of emotionally stable performance during a stress interview predict emotionally stable performance on the job (criterion-related validity)?

(2) Is the stress induced during the stress interview similar to the stress incurred on the job (content validity)?

(3) During a stressful interview, what are the behavioral indices of emotional stability?

(4) What would be the possible consequences of using the stress interview in terms of personal effects on the applicant and labor relations considerations?

In general, when deciding if it is appropriate to use stress interviews, consideration must be given to whether the tolerance to stress measured during the interview is actually related to tolerance to stress on the job. If not, the selection interview may not be valid and may be in violation of EEOC guidelins.

Nuclear facilities could also consider using a non-stress interview, that is, using either a structured, unstructured, or hybrid interview. It shoud be emphasized that while the selection interview has some ancillary benefits, as has been mentioned, its validity and reliability are somewhat questionable.

If an interview is to be used, nuclear facilities should consider the following points:

(1) structured interviews have higher inter-rater reliabilities than unstructured interviews.

(2) interviewers should be provided with a valid stereotype of a good worker.

(3) the questions asked and the traits measured should be job-related and valid.

(4) the questions asked should not be an invasion of the applicant's privacy.

6.8 Background and Reference Checks

6.8.1 Overview

Background checks refer to the verification of information received from, and about, job applicants. Information regarding applicants' military service, education, driving record, work history, medical history, police record, etc., can be verified through background checks. Determination can be made as to whether the applicant supplied false information during the application process.

Reference checks refer to obtaining character statements from the persons whose names have been provided to the organization by the applicant. Even though character references are generally favorable in nature, valuable information, verification of information, and secondary leads can be obtained from these references. (Ref. 130).

Although various methods exist for verifying and obtaining information about applicants, this document will focus primarily on the major approaches of letters of reference, contacting previous employers, and field investigations.

Letters of Reference - One method of verifying information is to request that the applicants provide the organization with letters of reference. This is accomplished either by requiring applicants to provide the letters or to supply the names of the persons who will submit letters. If names are given, the organization contacts these people and requests that they send letters of reference.

Contacting Previous Employers - Mosel and Goheen (Ref. 144) concluded that interviewing, by phone or personal contact, the applicant's past employment supervisors can be a useful source of information. This is accomplished by having the organization contact the previous supervisors and through a careful cross-check of opinions.

Field Investigations - Field investigations involve personal contact with the job applicant's previous employees, neighbors, friends, relatives, etc. Field investigations are generally performed by an experienced investigator.

Other methods that can be used include: contacting the National Personnel Records Center to verify military service record (Ref. 130), contacting state regulatory agencies to verify driving records, and contacting law enforcement agencies to verify conviction records.

6.8.2 Relevant Considerations

6.8.2.1 Reliability

With respect to background and reference checks, traditional evaluation has not specifically centered on test-retest, equivalent forms, internal consistency, or inter-rater reliability. In fact, very little research on the reliability of background checks has been conducted (Ref. 141). However, Mosel and Goheen (Ref. 144) indicated that of the research that has been conducted, it appears that the reliability of letters of reference is low.

6.8.2.2 Validity: Perspective

With respect to background and reference checks, little research has been carried out with regard to criterion-related, content, and construct validity.

Organizations using background and reference checks would find it advantageous to conduct criterion-related validity studies. By doing this, organizations could determine whether the information they were collecting was a valid predictor of on-the-job performance.

Organizatons should also conduct studies to determine whether falsification of information on an application is a valid predictor of future performance. Even though an applicant's driving record, per se, may not predict on-the-job performance, the fact that the applicant presented inaccurate information regarding his/her driving record may itself be a valid predictor.

Organizations may also find it advantageous to conduct content validity studies. In doing this, only information directly related to the content on the job would be collected in order to conform to EEOC guidelines.

6.8.2.3 Cost of Development and Administration

(1) Letters of Reference

The cost of developing the use of letters of reference is very minimal. No special forms or machinery are needed. All that must be done is to request applicants to provide their own letters of reference, or to provide the names and addresses of people who are willing to submit letters of reference. If the latter method is used, the organization can simply contact the people and request that reference letters be forwarded.

The cost of administration is directly related to the number of applicants and the salaries of those employees involved in the process. Other administrative costs, such as telephone, stationery, and mailing expenses, must be included, but are low.

(2) Contacting Previous Employers

The cost of developing the format and procedure for contacting previous employers is negligible. With respect to administration, the cost depends on the number of applicants and the salaries of those employees involved in the background checks. Other administrative costs, such as telephone, stationery, and mailing expenses, must be included but will most likely be low.

(3) Field Investigations

The cost of conducting a field investigation will be higher than that of letters of reference or contacting previous employers. An experienced investigator from outside the organization may need to be hired. Additionally, the expenses associated with conducting a field investigation (e.g., travel) may be considerable. Landy and Trumbo (Ref. 141) stated that a thorough field investigation may cost several hundred dollars, as compared to the cost of making a few telephone calls.

6.8.2.4 Administrator: Training and Qualifications

The person who performs the task of conducting the background and reference checks should be properly trained in this activity. If trained employees are not available, the organization should either hire outside experts (e. g., a trained field investigator) or train one or more of their own employees.

With regard to qualifications, administrators must be knowledgeable about the techniques and methods available for

collecting information. They must know where to obtain information about an applicant's military record, driving record, employment history, etc. Furthermore, if the method of collecting information includes personal contact with other persons, such as previous employers, the administrator should be skilled in the techniques of interviewing.

6.8.2.5 Administration of Background and Reference Checks: Time and Difficulty

(1) Letters of Reference

Letters of reference are very simple to administer. The administration merely involves requesting the applicant, or the persons designated by the applicant, to provide letters of reference about the applicant.

The time required for receiving and processing the letters may vary. If the applicant is required to provide the letters of reference, the letters may be available at the start of the application process. If the applicant provides the names of reference persons and the organization makes the requests, the letters may be available in a few days or a few weeks, depending on the referree's availability and promptness.

(2) Contacting Previous Employers

The process of contacting previous employers does not present major administrative difficulties. The organization can contact previous employers either by mail or telephone.

If contact is made via telephone, the information may be available immediately, that is, the persons to be contacted and the information to be gathered is immediately available and the information is conveyed directly over the phone. If the persons to be contacted are unavailable (e.g., out of town), or if the information regarding the applicant is not immediately accessible (e.g. it is in a file in another part of the building), it may be days before the information can be collected.

If contact is made via mail, the information will be available in approximately one week, at the earliest. The request for the information must be mailed to previous employers, the information must be gathered and prepared for mailing, and it must be mailed back to the requesting organization. Additionally, if the previous employers do not act on the request immediately, or the information is not readily accessible, it may take a considerable amount of time before the requesting organization receives the information.

(3) Field Investigations

The administration of field investigations is more difficult than the administration of letters of reference or contacting previous employers. Whereas the latter two approaches can generally be conducted from the organization's own office, this is not the case with field investigations, where a person must make personal contact with other designated persons.

Field investigations are generally more time consuming than either letters of reference or contacting previous employers. The persons to be contacted must be located and a time to meet with them must be arranged. Travel time may be considerable, especially if various persons live on opposite sides of the city, or in different cities, states, etc.

6.8.2.6 Equipment/Materials Needed for Administration

Field investigations may require additional equipment, such as transportation vehicles, which are not needed for the other methods of conducting background and reference checks. Conducting background and reference checks does not require any special or unique materials or equipment.

6.8.2.7 Personal Effects on Applicants

(1) Letters of Reference and Contacting Previous Employers

In general, obtaining letters of reference and contacting previous employers will not have any adverse effects on the applicants. The reason for this is that the applicants provide the names of the persons who will submit letters of reference and the names of the employers for whom they have worked. However, if an organization contacts an applicant's present employer without obtaining prior approval, the applicant may be adversely affected. The applicant may not have wished his/her present employer notified because he/she has not revealed his/her intention to change jobs.

(2) Field Investigations

In conducting field investigations, the investigator must be careful not to commit invasion of privacy violations. Applicants may feel that their privacy is being invaded if investigators ask neighbors and relatives questions about the applicant's personal non-work life. By ensuring that questions asked are job-related, applicants may feel more comfortable, but still may have concern over invasion of privacy considerations.

6.8.2.8 Compliance with Legal Issues and EEOC Guidelines

Obtaining letters of reference, contacting previous employers, and/or conducting field investigations may relate to legal and EEOC issues. That is, contacting present employers without the applicant's approval and/or conducting field investigations may be viewed as a violation of the applicant's right to privacy. Coupled with this is the need to demonstrate validity for the method used. However, the courts may determine that the information collected, even though it was validated, was not imperative enough to employee selection to allow an organization to violate an individual's right to privacy.

6.8.2.9 Confidentiality of Background and Reference Check Results

As with all other measurement techniques, organizations must concern themselves with controlling access to the results of background and reference checks. If the checks are conducted completely by the organization's employees, it is critical that the information is not further disseminated within the organization. If an outside investigating firm is employed, professional ethics should govern the control of the information.

6.8.2.10 Susceptibility to Faking

It is possible that the applicant's previous employers, and any other people contacted, may not provide the investigator with completely accurate information. The falsification of information may be either supportive or critical of the applicant.

One way to control for this is to obtain the information from multiple sources. This, in essence, would be establishing a type of inter-rater reliability. Another method would be to obtain only objective, verifiable information, such as military records, educational histories, and driving record.

6.8.2.11 Labor Relations Considerations

Unions and employees may not accept the methods of contacting present employers and conducting field investigations because of the invasion of privacy issue. Unionized organizations which are considering using these methods will find it to their advantage to confer with union representatives.

6.8.2.12 Reevaluation Considerations

If an applicant was refused employment because of information obtained through the letters of reference or the background checks, it may be appropriate to allow the individual to reapply at some later date. For example, if an applicant was denied employment because a previous employer reported that the applicant had an unacceptable attendance record, the applicant should be allowed to be reevaluated at some later date, as the applicant may be able to demonstrate a marked improvement during the course of an interim job.

6.8.2.13 Applicability of Background and Reference Checks to the Selection of Nuclear Facility Personnel with Regard to Emotional Instability

Conducting background checks for the purpose of verifying information provided by the applicants seems to be very appropriate in the context of selecting nuclear facility personnel. Of particular significance would be the credibility of applicants who falsify information. Since it is extremely important that nuclear facility personnel be reliable and trustworthy, it is doubtful whether applicants who falsify information would make reliable employees. However, in terms of emotional instability, information provided through these methods at best would be rather indirect measures of emotional instability.

Requiring letters of reference may not be necessary. Information received in these letters is generally favorable toward the applicant and thus doesn't validly distinguish among applicants.

6.9 Life Change Scales

6.9.1 Overview

Life change scales are a unique type of measurement technique which is designed to systematically identify stressful life events which have occurred with respect to one individual within a specified time frame. The objective of this technique is to derive a quantitative index of the total amount of stress incurred by the individual during this time period. This cumulative total obtained from the summation of numerical ratings assigned to the various events experienced, is used to predict the individual's chances of developing some type of illness in the near future. The development of this technique was based on some early research by H.G. Wolff (cited in Ref. 145) which suggested that stressful life events play a significant role in the development of disease by inducing neurophysiological reactions.

Although life change scales are administered in a format similar to that of the self-report personality inventories,

this technique is not intended to be a measure of personality traits nor can it be classified as a psychophysiological measurement technique. Its primary use is to serve as an adjunct source of information in identifying those individuals who have recently experienced excessive amounts of stress, and whose lifestyles, personality characteristics, and general states of health may compound the effects of this stress, resulting in a variety of mild to serious health problems.

There have been various research studies investigating life change scales, however, this review revealed only one such instrument which has been adequately researched and professionally developed. This specific instrument and the one most widely used, is the Social Readjustment Rating Scale (SRE) published by Holmes and Rahe in 1967 (Ref. 146). Therefore, this section will primarily address this specific instrument.

Holmes and Rahe conducted a great deal of the pioneering research aimed at relating stress to disease. Based on Wolff's early evidence, they conducted a systematic investigation of the relationship between social readjustment, stress, and susceptibility to illness. Their observations revealed repeated cases in which numerous stressful life events and/or unabated sources of stress requiring significant amounts of social readjustment preceded the onset of disease. These correlations, between stressful life events and disease onset, led to their attempt to develop a systematic method for predicting illness based on these life changes. The resulting SRE was developed following analysis of data compiled on more than 5,000 patients with a variety of medical disorders.

The SRE lists forty-three life events, each of which has a corresponding mean numerical value based on the ratings provided in the normative data. Those events with the higher numerical values (death of a spouse, divorce, marital separation, jail term, death of a close family member, and so on, in descending order) are considered the most stressful, and those with the lower values (e.g., change in eating habits, vacation, Christmas, etc.) are viewed as least stressful.

The SRE is self-administered following verbal instructions to the individual to check off those events which have happened to him/her during the past year. The corresponding values for each of these events are then summed, yielding one total numerical score. Holmes and Rahe (Ref. 146) defined a Life Crisis as a total score of 150 points or more accumulated during a 12-month period, and determined that this reflected roughly a 50% chance of developing an illness in the near future. A total score in excess of 300 points for a year raises the chances of onset of disease to almost 90%. As the score increases, so does the probability that the health change will be serious in nature (Ref. 145).

It is interesting to note that although the SRE is designed to predict onset of illness, no specific disorders are predicted. This is because, although a number of individuals may be subjected to the same stressful events, each individual's unique personality configuration and psychophysiological profile appears to heavily contribute to the specific organ system(s) affected, as well as the nature of the disorder (Ref. 145).

Paykel, Prusoff, and Uhlenhuth (Ref. 147) conducted a study in which they modified the SRE format in an attempt to both replicate the basic findings of Holmes and Rahe (Ref. 146), as well as to extend the utility of this technique to potential prediction of psychiatric disturbances and emotional instability. Rather than utilizing the concept of "social readjustment," these authors focused on the subjective experience of distress or being "upset" by the 61 life events included on their scale. These events were additions, deletions, and variations in wording of the items listed on the SRE. A sample of 373 subjects were asked to rate all of the events included on the scale in terms of how "upsetting" they would be if they were to occur to themselves personally. The rating scale used was a 0-20 equal interval range, with 0 representing "least upsetting" and 20 representing "most upsetting". Thus, the events were not assigned disproportionate weights.

After rating each of the events, the subjects were then instructed to go back and make an "X" next to each event which they had actually experienced during the past year. Just as Holmes and Rahe (Ref. 146) had utilized medical patients as subjects in developing the SRE, the sample employed by Paykel et.al. (Ref. 147) included 213 psychiatric patients and 160 relatives of the patients. Results of this study revealed that events rated as most upsetting by psychiatric patients often included threats rather than actual occurrences, traumatic events requiring little actual life change (such as death of a parent in a nursing home) and blows to self-esteem. The events rated as least upsetting included those usually regarded as pleasurable and those requiring minimal readjustment. Since many of the events differed from those on the SRE, this study compared the ratings on the fourteen items which were identical to items included by Holmes and Rahe. The mean correlation between these items on the two scales was .68. It appears reasonable that even higher correlations would be obtained if a greater number of items using the same wording were compared for the medical and psychiatric patients. Despite the inconclusive results of this study regarding whether psychiatric disturbances can be predicted with instruments similar to the SRE, it appears to represent an important area for further research relating life change scales to the prediction of emotional disturbances.

6.9.2 Relevant Considerations

6.9.2.1 Reliability

With respect to the SRE, traditional evaluation studies have centered on the inter-rater reliability approach to the measurement of reliability. The equivalent forms reliability approach has been used in studies employing modifications of the SRE, such as the forms developed by Paykel, et.al. (Ref. 147). Since these modifications do not represent systematically developed equivalent forms with regard to the SRE, however, the equivalent forms reliability coefficients provided by these authors cannot be justifiably applied to reliability measurement of the Standard SRE.

The inter-rater reliability approach was utilized by Holmes and Rahe (Ref. 146) in the construction of the SRE. On the basis of their preliminary research investigating thousands of patients and the life events preceding illness onset, they isolated forty-three events which appeared to be representative of life changes requiring the greatest magnitude of social readjustment. The authors defined social readjustment as the degree and duration of a person's accommodation to a life event, regardless of the event's desirability. A list of these forty-three events was then given to 394 subjects who were asked to assign each event a magnitude value in terms of the degree of readjustment required to cope with the event. Ratings were based on ratios in comparison to the event of "marriage", which was assigned a fixed value by the authors. Holmes and Rahe (Ref. 146) reported substantial agreement across different age, social, cultural and ethnic groups, in regard to the magnitude ratings of the forty-three events. The mean values of the reported magnitudes for each event were then assigned to these items in descending order, resulting in the present version of the SRE. Thus, the development of this scale was based on an inter-rater approach to test construction. This review did not reveal any subsequent studies using the inter-rater reliability approach in evaluating the SRE.

6.9.2.2 Validity

With respect to the SRE, traditional evaluation studies have centered on the criterion-related approach to the measurement of validity.

Rahe (Ref. 148) conducted a study of 2,500 officers and enlisted men stationed aboard three Naval Ships. The SRE was administered to the total sample just prior to the beginning of a cruise. Of the 2,500 subjects, the 30% with the highest SRE scores contracted almost 90% more illnesses during the first month of the cruise than the 30% with the lowest scores. Throughout the remainder of the cruise, the high-scoring group consistently developed more illnesses that the low-scoring groups.

In a study conducted by Holmes and Masuda (Ref. 149), 84 physicians were administered the SRE in standard format with the exception of including events from the previous 18 months rather than 12 months. Data regarding disease incidence were gathered eight months later. The result revealed that 49% of the group with SRE scores of 300 or greater had developed illnesses, 25% of the group with scores ranging from 200-299 had reported illnesses, and only 9% of the groups with SRE scores from 150 to 199 had become ill.

The results of these investigations suggest that the SRE does demonstrate significant predictive validity.

6.9.2.3 Cost of Development and Administration

The SRE forms are not available through any test publishing organizations. Subsequent to the initial journal publication of the scale, however, the SRE was copyrighted by Pergamon Press, Inc., in 1967. The test may be reproduced with permission from Pergamon Press for a reproduction permission fee of $25.00. Upon payment of this fee and receipt of documented reproduction permission, the SRE may be reproduced in unlimited quantities by the organization granted reproduction rights.

6.9.2.4 Administrator: Training and Qualifications

There is no specific training required for administration of the SRE, since it is a self- administered scale. The administrator should take responsibility for clarifying the instructions to the individual and for computing the sum of the values for life events checked.

There are no detailed interpretations necessary for the SRE. However, if included as part of a selection procedure, the results of this scale should be submitted to the individual(s) responsible for interpreting the other measurement techniques included in the selection procedure. At this point, these data could be incorporated into the evaluation of an applicant's suitability for hiring.

6.9.2.5 Administration of Measurement Technique: Time and Difficulty

Assuming that the administrator takes responsibility for calculating the sum of the values corresponding to the forty three events listed, the form should be easily completed in ten to fifteen minutes. Exceptions to this rule might be expected in individuals demonstrating difficulty in recalling when certain events occurred (i.e., whether they were included in the preceding twelve months or prior to this time frame).

6.9.2.6 Equipment/Materials Needed for Administration

The only materials needed for administration of the SRE are the form itself and a pen or pencil.

6.9.2.7 Personal Effects on Applicants

Because the SRE is an index of stressful life events, certain individuals might experience some degree of emotional upset upon being asked to recall events that were significantly distressing, particularly if they occurred recently prior to administration of the SRE. The possibility of such emotional upset appears to be the only potentially detrimental effect.

6.9.2.8 Compliance with Legal Issues and EEOC Guidelines

The SRE appears to demonstrate adequate criterion-related validity with regard to predicting onset of illness, however the question remains as to how and whether this criterion relates to emotional instability and job performance.

Holmes and Rahe (Ref. 146) reported that there was substantial agreement on SRE ratings across various age, social, cultural and ethnic groups. Follow-up studies, however, revealed significant differences in ratings among white, black, and Mexican subjects (Ref. 150) and between American and Japanese subjects (Ref. 151). Such findings are not surprising, given the unique nature of this instrument, in that it lists events which might be expected to reflect variations in cross-cultural values and customs. Because of these data, however, compliance of the SRE with EEOC guidelines remains questionable at this point.

6.9.2.9 Confidentiality of Measurement Technique Results

Although the SRE does not technically come under the realm of psychological tests, it reflects information of a personal nature. Therefore, scores from the SRE should be released only to designated individuals who would be responsible for personnel screening and selection. Exceptions to this rule could result in legal ramifications regarding invasion of privacy.

6.9.2.10 Susceptibility to Faking

There are no elements of the SRE designed to detect distorted responses. It is therefore quite possible that an applicant might elect not to check certain events which he/she has, in fact, experienced during the past year. Events on the SRE such as "jail term" or "fired at work" would be good examples of events that might not be checked. The only apparent way to detect such distortion would be to compare the events checked on the SRE with data yielded by other techniques such as background investigations and pre-employment interviews.

6.9.2.11 Labor Relations Considerations

This review did not reveal any labor relations disputes which centered on the use of the SRE. The possibility of such action cannot be ruled out, however, if legal or ethical violations relating to the use of this scale were demonstrated.

6.9.2.12 Reevaluation Considerations

In accordance with EEOC guidelines, a job applicant screened out on the basis of test results should be provided with the opportunity for reevaluation. This option appears particularly appropriate with regard to the SRE, since administration at a later date would change the time frame for inclusion of subsequent events.

If the SRE were to be included as part of a selection procedure, it would appear advisable to administer periodic reevaluations of those already hired, as part of a routine check on emotional stability and as a means of identifying employees' changes. This type of information could be helpful to employers who might consider offering such individuals the option of receiving personal counseling or temporary reduction in job duties, or a leave of absence, etc.

6.9.2.13 Applicability of the Social Readjustment Rating Scale to Selection of Nuclear Facility Personnel with Regard to Emotional Instability

The importance of having a method for predicting physical illness is significant in itself in terms of employment selection, because regardless of an individual's level of performance and proficiency, significant health changes are likely to reduce his/her productivity or even necessitate his/her leaving his/her position. Of even greater significance for the purposes of this review, are the research findings based on the SRE which suggest that the onset of psychiatric difficulties can also be predicted using this type of instrument.

Although substantial further research is needed, it appears reasonable, based on the preceding discussion, that the SRE or modifications of this instrument might be useful in predicting emotional as well as physical disorders.

6.10 Zero Input Tracking Analyzer/Auxiliary Distraction Task (ZITA/ADT)

6.10.1 Overview

The Zero Input Tracking Analyzer/Auxiliary Distraction Task (ZITA/ADT) system was developed by Norman Walker and Associates (Ref. 152). This system was originally developed to

predict the performance of military missile system operators under stressful conditions, particularly those associated with the stress of combat.

Grubinich (Ref. 153) outlined the basic operation of the ZITA instrument as follows: The subject sits facing an oscilloscope which has a moving dot of light on the screen. He/She is instructed to use the two-position control stick to keep the dot as close to the center of the screen as possible. The speed and the lateral movement of the dot can both be increased or decreased to alter task difficulty. The ADT component is then turned on to introduce stressful conditions. This is accomplished by the generation of a series of bimodal pulses at 1, 2, 4 or 8 second time intervals. The subject has to respond to these signals by pressing the appropriate one of two buttons, while continuing to use the other hand to keep the dot in the center of the screen. This stressor is consistent with Walker's (Ref. 152) definition of psychological stress as including an overload of external, unwanted information to which the person must respond. Walker feels that this factor is the primary cause of performance decrements. Throughout the administration of the ZITA/ADT, errors are recorded for presentation on both a digital display and a graphic representation of the oscilloscope screen.

Since the initial development of the ZITA/ADT system, its applications have been extended for use in a number of areas including education, armed forces, medicine, personnel selection and training, and the criminal justice system.

6.10.2 Relevant Considerations

6.10.2.1 Reliability

With respect to the ZITA/ADT, traditional evaluation studies have centered on the internal consistency reliability approach to the measurement of reliability.

Walker and Walker (Ref. 154) reported on the internal consistency reliability for two ZITA tasks of varying levels of task difficulty. For each of these tasks, the reliability coefficients were derived by correlating each subject's mean score for the entire task with his/her scores on the various individual phases of the task. For Task 2, which is an acceleration control task, reliability coefficients ranged from .57 during the training phase of the run to .87 during the final phase with a mean of .73. Task 3, which is more difficult than Task 2 and involves a time delay, revealed correlation coefficients ranging from .50 to .80, with a mean of approximately .68. No data regarding Task 1 were available at the time of this review.

6.10.2.2 Validity

With respect to the ZITA/ADT, traditional evaluation studies have focused on the criterion-related approach to the measurement of validity.

Walker (Ref. 152) reported that mean results from the ZITA test could provide good estimates of the accuracy of military control systems during peacetime operation. Additionally, he noted that the ADT stressor produced the same relative effects on ZITA performance accuracy as those derived from combat data. Precise data from these investigations, including clear specifications of criteria used, were unavailable at the time of this review. Walker (Ref. 152) reported that similar relationships between ZITA/ADT data and variations in performance were also seen in groups of fighter pilots as well as in civilian test data. One of these studies (Ref. 152) comparing performance of keypunch operators with their composite ZITA scores revealed a correlation coefficient of .96 between composite ZITA scores and keypunches per hour.

6.10.2.3 Cost of Development and Administration

The price of purchasing the ZITA/ADT system is approximately $12,000.* If purchase of the instrument itself is not desired, testing services using the ZITA/ADT are available through N. K. Walker Associates, Inc. (Ref. 152) at a cost of $300 per subject tested, plus any travel expenses incurred.

6.10.2.4 Administrator: Training and Qualifications

Administration of the ZITA/ADT system is fairly easily learned, although the administrator must become familiar with both the equipment and the instructions to be given to the subjects. This training can generally be provided in less than a day*. Interpretation of ZITA/ADT data is a more complex undertaking, however, requiring analysis of the test data and comparison to available data bases. Thus, although this review did not reveal any specific qualifications necessary for ZITA/ADT interpretation, it would appear that considerable expertise is vital in this regard.

* N. Walker, Personal Communication, June 6 (1980).

6.10.2.5 Administration of ZITA/ADT: Time and Difficulty

Grubinich (Ref. 153) noted that the complete ZITA/ADT test, including practice sessions and two trials on both the stressed and unstressed segments, can be completed in less than ten minutes. Assuming that the subject has full use of both arms and hands, there appears to be no inherent difficulties associated with ZITA/ADT administration.

6.10.2.6 Equipment/Materials Needed for Administration

Both the ZITA and the ADT instruments are fully contained in a 28-pound suitcase. This kit also contains computation capabilities as well as the mirror and pen recorder used to record errors.

6.10.2.7 Personal Effects on Applicants

It is conceivable that some individuals might experience some degree of frustration if unable to control errors in performing the tracking task.* In addition, negative psychological effects could be incurred by ZITA/ADT administration if an applicant were screened out on the basis of data generated by this instrument. Aside from these possibilities, there are no apparent detrimental personal effects associated with ZITA/ADT administration.

6.10.2.8 Compliance with Legal Issues and EEOC Guidelines

This review did not reveal any court cases or legal issues which centered on the use of the ZITA/ADT. Additionally, there were no obvious violations of EEOC guidelines noted. The possibility of such cases arising in the future cannot be ruled out, however, particularly if the ZITA/ADT system was found to be discriminatory in nature.

* This would be especially likely if poor performance were due to psychomotor deficits, poor eye-hand coordination, or poor visual tracking abilities, rather than low stress tolerance.

6.10.2.9 Confidentiality of ZITA/ADT Results

As with data generated by other selection instruments, all results of ZITA/ADT testing should be kept confidential with only those individuals clearly associated with the selection process having access to these data.

6.10.2.10 Susceptibility to Faking

Given the nature of the ZITA/ADT instrument and tasks, it appears that the only potential for faking on this test would be associated with faking intentionally inferior performance. Such a tendency is unlikely to occur, however, in an employment application and screening procedure. Thus, this instrument appears to be minimally susceptible to faking.

6.10.2.11 Labor Relations Considerations

This review did not reveal any labor disputes which centered on the use of the ZITA/ADT. The possibility of such conflicts in the future cannot be ruled out, however, and might occur if this instrument were determined to be discriminatory or violate rights to privacy.

6.10.2.12 Reevaluation Considerations

In accordance with EEOC guidelines, an applicant screened out on the basis of ZITA/ADT results should have the opportunity for reevaluation at a later date. It also appears reasonable that if this instrument were used as part of a selection procedure, it might be useful as a periodic re-check on resistance to stress in those already hired.

6.10.2.13 Applicability of the ZITA/ADT to Selection of Nuclear Facility Personnel with Regard to Emotional Instability

Based on available information, the ZITA/ADT system appears to be a potentially useful instrument for predicting a job applicant's performance under stress, particularly stress due to information overload. Since panic reactions to crisis situations as well as reactions to the accumulated effects of stress have both been delineated as indices of emotional instability by this review, a technique designed to measure these tendencies seems to warrant consideration. Caution must be taken, however, to assure that factors other than panic reactions to stress have not been the cause of poor performance. Before this instrument could justifiably be included in a selection procedure for nuclear facility personnel, however, further research designed to validate the ZITA/ADT would appear advisable with respect to the criteria of emotional instability associated with focal personnel positions.

6.11 References - Section 6

1. A. Anastasi, Psychological Testing, MacMillan Publishing Company. New York, 1976.

2. J.B. Rotter and D.J. Hochreich, Personality, Scott, Foresman and Company. Glenview, Illinois, 1975.

3. S.R. Hathaway and J.C. McKinley, Minnesota Multiphasic Personality Inventory Manual, The Psychological Corporation. New York, 1967.

4. S.R. Hathaway, "A Coding System for MMPI Profiles," Journal of Consulting Psychology II, 334-337 (1974). Available in public technical libraries.

5. G.S. Welsh, "An Extension of Hathaway's MMPI Profile Coding System," Journal of Consulting Psychology 12, 343-344 (1948). Available in public technical libraries.

6. C. Spielberger, Police Selection and Evaluation: Issues and Techniques, Hemisphere Publishing Corporation. Washington, D.C., 1979.

7. American Psychological Association, Ethical Standards of Psychologists (Rev. ed.), APA. Washington, D.C., 1977. Available from APA.

8. L.E. Tyler, Tests and Measurements, Prentice-Hall, Inc. Englewood Cliffs, NJ, 1971.

9. H.G. Gough, Manual for the California Psychological Inventory, Consulting Psychologists Press. Palo Alto, California, 1975.

10. T.J. Bouchard, "Review of the 16PF," in The Seventh Mental Measurements Yearbook (Vol. 1), O.K. Buros, Ed. (The Gryphon Press, Highland Park, NJ, 1972).

11. L.V. Gordon, Manual for the Gordon Personal Profile-Inventory (Rev. ed.), The Psychological Corporation. New York, 1978.

12. J.R. Braun, S. Alexander, and R. Weiss, "Relationship Between the Gordon Personal Inventory and Consensus of Peer Ratings," Psychological Reports 9, 455 (1961). Available in public technical libraries.

13. H.J. Eysenck and S.B.G. Eysenck, Manual for the Eysenck Personality Inventory, Educational and Industrial Testing Service. San Diego, California, 1968.

14. J.P. Guilford and W.S. Zimmerman, The Guilford- Zimmerman Temperament Survey, Sheridan Supply Company. Beverly Hills, California, 1955.

15. H.J. Eysenck, The Manual of the Maudsley Personality Inventory, Educational and Industrial Testing Service. San Diego, California, 1962.

16. H.J. Eysenck and S.B.G. Eysenck, The Manual for the Eysenck Personality Questionnaire, Educational and Industrial Testing Service. San Diego, California, 1975.

17. R.B. Cattell, IPAT Anxiety Scale, Institute for Personality and Ability Testings. Champaign, Illinois, 1957(a).

18. I.H. Scheier and R.B. Cattell, Neuroticism Scale Questionnaire, Institute for Personality and Ability Testing. Champaign, Illinois, 1957.

19. M. Zuckerman and B. Lubin, The Multiple Affect Adjective Checklist, Educational and Industrial Testing Service. San Diego, California, 1965.

20. H.G. Gough, California Psychological Inventory, Consulting Psychologists Press. Palo Alto, California, 1957.

21. J. Cannon, The Wisdom of the Body, Norton. New York, 1932.

22. H. Selye, The Stress of Life, McGraw-Hill. New York, 1956.

23. R.M. Yerkes and J.D. Dodson, "The Relation of Strength of Stimulus to Rapidity of Habit Formation," Journal of Comparative and Neurological Psychology 18, 459 (1908). Available in public technical libraries.

24. R.B. Malmo, "Activation: A Neurophysiological Dimension," Psychological Review 66, 367-386 (1959). Available in public technical libraries.

25. H. Sjobert, "Interaction of Task Difficulty, Activation, and Work Load," Journal of Human Stress, 33-38 (1977). Available in public technical libraries.

26. G.S. Krahebuhl, J.R. Marret, and G.B. Reid, "Stress and Simulation in Pilot Training," JSAS Catalog of Selected Documents in Psychology 9, 74 (1979), (MS. No. 1887) Available in public technical libraries.

27. B.B. Weybrew, "Pattern of Physiological Response to Military Stress," in Psychological Stress, M. Appley and R. Trumbull, Eds (Appleton-Century-Crofts, New York, 1967).

28. G.H. Lawrence and L.C. Johnson, "Biofeedback and Performance," in Psychological Stress, M. Appley and R. Trumbull, Eds. (Appleton-Century-Crofts, New York, 1967).

29. D.W. Shearn, "Operant Conditioning of Heart Rate," Science 137, 530-531 (1962). Available in public technical libraries.

30. T.W. Frazier, "Avoidance Conditioning of Heart Rate in Humans," Psychophysiology 3, 188-202 (1966). Available in public technical libraries.

31. H.D. Kimmel and F.A. Hill, "Operant Conditioning of the GSR, Psychological Reports 7, 555-562 (1960). Available in public technical libraries.

32. W.A. Greene, "Operant Conditioning of the GSR Using Partial Reinforcement," Psychological Reports 19, 571-578 (1976). Available in public technical libraries.

33. N.E. Miller, "Learning of Visceral and Glandular Responses," Science 163, 434-445 (1969). Available in public technical libraries.

34. "Conscious Control of Brain Waves," pp. 57-60, Psychology Today, 1, 1968. Available in public libraries.

35. E.B. Blanchard and L.H. Epstein, A Biofeedback Primer, Addison-Wesley. Reading, Mass., 1978.

36. K.R. Gaarder and P.S. Montgomery, Clinical Biofeedback: A Procedural Manual, Williams and Wilkins. Baltimore, 1977.

37. J. Orlansky, "An Assessment of Lie Detection Capability," JSAS Catalog of Selected Documents in Psychology 4, 104 (1974). (Ms. No. 728) Available in public technical libraries.

38. G.H. Barland and D.C. Raskin, "An Evaluation of Field Techniques in Detection of Deception," Psychophysiology 12, 321-330 (1975). Available in public technical libraries.

39. J.E. Reid, "A Revised Questioning Technique in Lie Detection Tests," Journal of Criminal Law and Criminology 37, 542-547 (1947). Available in public technical libraries.

40. J.A. Belt and P.B. Holden, "Polygraph Usage Among Major U.S. Corporations," Personnel Journal 57, 80-86 (1978). Available in public technical libraries.

41. K.S. Balloun and D.S. Holmes, "Effects of Repeated Examinations on the Ability to Detect Guilt with a Polygraphic Examination: A Laboratory Experiment with a Real Crime," Journal of Applied Psychology 64, 316-322 (1979). Available in public technical libraries.

42. D.T. Lykken, A Tremor in the Blood: Uses and Abuses of the Lie Detector, McGraw-Hill. New York, 1980.

43. J.A. Podlesny and D.C. Raskin, "Effectiveness of Techniques and Physiological Measures in the Detection of Deception," Psychophysiology 15, 344-359 (1978). Available in public technical libraries.

44. D.C. Raskin and R.D. Hare, "Psychopathy and Detection of Deception in a Prison Population," Psychophysiology 15, 126-136 (1978). Available in public technical libraries.

45. D.T. Lykken, "The Psychopath and the Lie Detector," Psychophysiology 15, 137-142 (1978). Available in public technical libraries.

46. H. Sandler, E.P. McCutcheon, T.B. Fryer, S. Rositano, R. Westbrook, and P. Haro, "Recent NASA Contributions to Biomedical Telemetry," American Psychologist 30, 257-264 (1975). Available in public technical libraries.

47. C.E. Williams and K.N. Stevens, "On Determining the Emotional State of Pilots During Flight: An Explanatory Study," Aerospace Medicine 40, 1369-1372 (1969). Available in public technical libraries.

48. G.A. Smith, "Voice Analysis for the Measurement of Anxiety," British Journal of Medical Psychology 50, 367-373 (1977). Available in public technical libraries.

49. F. Horvath, 'Effect of Different Motivational Instructions on Detection of Deception with the Psychological Stress Evaluator and the Galvanic Skin Response," Journal of Applied Psychology 64, 323-330 (1979). Available in public technical libraries.

50. J. Montgomery, "The Psychological Stress Evaluator in Truth Verification," Assets Protection 3, 6-13 (1978). Available in public technical libraries.

51. R.B. Finkle, "Managerial Assessment Centers," in Handbook of Industrial and Organizational Psychology, M.D. Dunnette, Ed. (Rand McNally College Publishing Company, Chicago, 1976).

52. D.W. Bray, R.J. Campbell, and D.L. Grant, Formative Years In Business: A Long-Term AT&T Study of Managerial Lives, John Wiley and Sons, Inc. New York, 1974.

53. D.W. Bray and D.L. Grant, "The Assessment Center in the Measurement for Potential in Business Management," Psychological Monographs 80 (1966). Available in public technical libraries.

54. J.R. Huck, "Assessment Centers: A Review of the External and Internal Validities," Personnel Psychology 26, 192-212 (1973). Available in public technical libraries.

55. C.L. Jaffee, "Assessment Center Help Find Management Potential," Bell Telephone Magazine 44, 18-25 (1965). Available in public libraries.

56. Standards and Ethical Considerations for Assessment Center Operations. Journal of Assessment Center Technology 2, 19-23 (1979). Available in public technical libraries.

57. S.L. Cohen, "Standardization of Assessment Center Technology: Some Critical Concerns," Journal of Assessment Center Technology 1, 1-10 (1978). Available in public technical libraries.

58. J.L. Moses, "The Development of An Assessment Center for the Early Identification of Supervisory Potential," Personnel Psychology 26, 569-580 (1973). Available in public libraries.

59. J.H. McConnell and T. Parker, "An Assessment Center Program for Multiorganizational Use," Training and Development Journal 26, 6-14 (1972). Available in public technical libraries.

60. D.J. Archambeau, "Relationships Among Skill Ratings Assigned in an Assessment Center," Journal of Assessment Center Technology 2, 7-20 (1979). Available in public technical libraries.

61. J.R. Hinrichs and S. Haanpera, "Reliability of Measurement in Situational Exercises: An Assessment of the Assessment Center Method," Personnel Psychology, 31-40 (1976). Available in public technical libraries.

62. C.F. Dicken and J.D. Black, "Predictive Validity of Psychometric Evaluations of Supervisors," Journal of Applied Psychology 49, 34-37 (1965). Available in public technical libraries.

63. A. Howard, "An Assessment of Assessment Centers," Academy of Management Journal 17, 115-134 (1974). Available in public technical libraries.

64. N. Schmitt, "Interrater Agreement on Dimensionality and Combination of Assessment Center Judgments," Journal of Applied Psychology 62, 171-176 (1977). Available in public technical libraries.

65. H.C. Thomson, "Comparison on Predictor and Criterion Judgments of Managerial Performance Using the Multitrait-Multimethod Approach," Journal of Applied Psychology 54, 496-502 (1970). Available in public technical libraries.

66. D.L. Denning and D.L. Grant, "Knowledge of the Assessment Process: Does it Influence Candidate Ratings?," Journal of Assessment Center Technology 2, 7-12 (1979). Available in public technical libraries.

67. F.O. Carleton, "Relationships Between Follow-up Evaluations and Information Developed in a Management Assessment Center," in Proceedings of the Annual Convention of the American Psychological Association, Miami Beach, 1970. Available from APA.

68. R.M. Finley, "An Evaluation of Behavior Predictions from Projective Tests in a Management Assessment Center," in Proceedings of the 78th Annual Convention of the American Psychological Association, Miami Beach, 1970. Available from APA.

69. J.M. Greenwood and W.J. McNamara, "Interrater Reliability in Situational Tests. Journal of Applied Psychology 51, 101-106 (1967). Available in public technical libraries.

70. S.A. Richards and C.L. Jaffee, "Blacks Supervising Whites: A Study of Interracial Difficulties in Working Together In a Simulated Organization," Journal of Applied Psychology 56, 234-240 (1972). Available in public technical libraries.

71. American Telephone and Telegraph, Personnel Research Staff. Personnel Assessment Program: Follow-Up Study, New York: AT&T, 1965.

72. D.W. Bray and R.J. Campbell, "Selection of Salesmen by Means of an Assessment Center," Journal of Applied Psychology 52, 36-41 (1968). Available in public technical libraries.

73. J.R. Huck and D.W. Bray, "Management Assessment Center Evaluations and Subsequent Job Performance of White and Black Females," Personnel Psychology 29, 13-30 (1976). Available in public technical libraries.

74. J.L. Moses and U.R. Boehm, "Relationship of Assessment Center Performance to Management Progress of Women," Journal of Applied Psychology 60, 527-529 (1975). Available in public technical libraries.

75. J.L. Moses and S.J. Wall, "Pre-Hire Assessment: A Validity Study of A New Approach for Hiring College Graduates," Assessment and Development Newsletter 2, Development Dimensions Press, 1975.

76. J.R. Hinrichs, "Comparison of Real Life Assessments of Management Potential with Situational Exercises, Paper-and-Pencil Ability Tests, and Personality Inventories," Journal of Applied Psychology 53, 425-433 (1969). Available in public technical libraries.

77. A.I. Kraut and C.J. Scott, "Validity of An Operational Management Assessment Program," Journal of Applied Psychology 56, 124-129 (1972). Available in public technical libraries.

78. H.B. Wollowick and W.J. McNamara, "Relationship of the Components of An Assessment Center to Management Success," Journal of Applied Psychology 53, 348-352 (1969). Available in public technical libraries.

79. H.H. Meyer, "The Validity of the In-Basket as a Measure of Managerial Performance," Personnel Psychology 23, 297-301 (1970). Available in public libraries.

80. V.J. Bentz, "The Sear's Experience in the Investigation, Description and Prediction of Executive Behavior," in Measuring Executive Performance, F.R. Wickert and D.E. McFarland, Eds. (Appleton-Century-Crofts, New York, 1967).

81. J.R. Huck, "The Research Base," in Applying the Assessment Center Method, J.L. Moses and W.C. Byham, Eds. (Pergamon Press, New York, 1977).

82. Wickes Corporation. An evaluation of the management indoctrination program. Management Evaluation Department, Saginaw, Michigan, 1974. Available from Wickes Corporation.

83. W.C. Byham, "Assessment Centers for Spotting Future Managers," Harvard Business Review, 150-160 (1970). Available in public libraries.

84. B.M. Cohen, J.L. Moses, and W.C. Byham, The Validity of Assessment Centers: A Literature Review (Monograph II), Development Dimensions Press, Pittsburgh, 1974.

85. K. McNutt, "Behavioral Consistency and Assessment Centers: A Reconciliation of the Literature," Journal of Assessment Center Technology 2, 1-6 (1979). Available in public technical libraries.

86. R.D. Neidig, J.C. Martin and R.E. Yates, "The Contribution of Exercise Skill Ratings to Final Assessment Center Evaluations," Journal of Assessment Center Technology 2, 21-23 (1979). Available in public technical libraries.

87. L. Sands and J.E. Harper, "The Use of Assessment Center Techniques in the Selection of Five Department Personnel: A Unique Application," Journal of Assessment Center Technology 2, 10-18 (1979). Available in public technical libraries.

88. G.P. Tielsch and P.M. Whisenand, The Assessment Center Approach in the Selection of Police Personnel, Davis Publishing Co., Inc. Santa Cruz, California, 1977.

89. S.L. Cohen, "The Bottom Line on Assessment Center Technology," Personnel Administrator 25, 50-56 (1980). Available in public technical libraries.

90. M.D. McCrimmon, "Strategies to Increase the Cost Effectiveness of the Assessment Center Method," Journal of Assessment Center Technology 1, 12-16 (1978). Available in public technical libraries.

91. R.J. Filer, "The Assessment Center Method in the Selection of Law-Enforcement Officers," in Police Selection Evaluation, C.D. Spielberger, Ed. (Hemisphere Publishing Corporation, Washington D.C., 1979).

92. S.L. Cohen, "Standardization of Assessment Center Technology: Some Critical Concerns," Journal of Assessment Center Technology 1, 1-10 (1978). Available in public technical libraries.

93. F.D. Frank and D. Whipple, "An Assessor Certification Program: Based on Simulation of The Assessor Job," Journal of Assessment Center Technology 1, 1-4 (1978). Available in public technical libraries.

94. W.L. Dodd, "Attitudes Toward Assessment Center Programs," in Applying the Assessment Center Method, J.L. Moses and W.C. Byham, Eds. (Pergamon Press, New York, 1977).

95. D.J. Nirtaut, "Assessment Centers: An Examination of Participant Reaction and Adverse Effects," Journal of Assessment Center Technology 1, 18-23 (1978). Available in public technical libraries.

96. L.W. Slivinski, U.S. McDonald, and R.P. Bourgeois, "Immediate and Long-Term Reactions to an Assessment Centre," Journal of Assessment Center Technology 2, 13-18 (1979). Available in public technical libraries.

97. M. Quaintance, "A Unique Feedback Approach to Unsuccessful Executive Candidates," Journal of Assessment Center Technology 1, 11-14 (1978). Available in public technical libraries.

98. First Court Test of Assessment Center Method, Assessment and Development Newsletter, Development Dimensions, Inc. 3, 1-2 (1976). Available from DDI.

99. M.C. Mendenhall, "A Report of an Assessment Center for the Position of Deputy Police Chief," Personnel Department, Omaha, Nebraska 1977. Available from Personnel Department, City of Omaha.

100. G.C. Thorton III, "Varieties of Validity of Assessment Centers in Other Companies," in Validity of Assessment Centers, W.C. Byham (Chair). In Proceedings of the 79th Annual Convention of the American Psychological Association, Washington, D.C., 1971. Available from APA.

101. United States District Court, Eastern District of Virginia, Richmond Division. The Richmond Black Police Officers Association, et. al., vs. The City of Richmond, et. al. Civil Action No. 74-0267-R, September 22,1975. Available in public technical libraries.

102. D.F. Hoyle, "Complete Assessment of Nearly 1700 Women Under Consent Agreement," Assessment and Development 2, 4-5 (1975). Available in public technical libraries.

103. United States Office of Personnel Management. Firefighters vs. City of St. Louis. Equal Employment Opportunity Court Cases, 1979, September, OIPP. Available in public technical libraries.

104. G.A. Brugnoli, J.E. Campion, and J.A. Basen, "Racial Bias In the Use of Work Samples For Personnel Selection," Journal of Applied Psychology 64, 119-123 (1979). Available in public technical libraries.

105. P.D. Frank and R.J. Drucker, "The Influence of Evaluatee's Sex on Evaluation of A Response on a Managerial Selection Instrument," Sex Roles: A Journal of Research (1977). Available in public technical libraries.

106. B.M. Bass, "The Leaderless Group Discussion," Psychological Bulletin," 51, 465-492 (1954). Available in public technical libraries.

107. M.M. Petty, "A Multivariate Analysis of the Effects of Experience and Training Upon Performance In a Leaderless Group Discussion," Personnel Psychology 27, 271-282 (1974). Available in public technical libraries.

108. C.L. Jaffee and C.E. Michaels, "Is In-Basket Performance Subject to Coaching Effects?" Journal of Assessment Center Technology 1, 13-17 (1978). Available in public technical libraries.

109. M. Struth and F. Frank, "The Effects of Assessor Training on Subsequent Performance in an Assessment Center," in Proceedings of the International Personnel Management Association/Assessment Council, July 1980. Available for purchase from National Technical Information Service, Springfield, Virginia 22161.

110. F.D. Frank and C. Tollon, "Use of Assessment Centers for Selection of Entry Level Police Officers," in Proceedings of the International Congress on Assessment Centers, New Orleans, 1979. Available for purchase from National Technical Information Service, Springfield, Virginia 22161.

111. W.A. Owens, "Background Data," in Handbook of Industrial and Organizational Psychology, M.D. Dunette, Ed. (Rand McNally College Publishing Company, Chicago, 1976).

112. Institute for Behavioral Research In Creativity (IBRIC), Manual for Alpha Biographical Inventory. Predictions Press, Greensboro, N.C., 1968.

113. F.B. Chaney and W.A. Owens, "Life History Antecedents of Sales Research, and General Engineering Interest," Journal of Applied Psychology 48, 101-105 (1964). Available in public technical libraries.

114. M. Baehr and G.B. Williams, "Underlying Dimensions of Personnel Background Data and Their Relationship to Occupational Classification," Journal of Applied Psychology 51, 481-490 (1967). Available in public technical libraries.

115. J.P. Guilford and J.L. Lacey, Eds. Printed classification tests. AAF Aviation Psychology Research Program Reports. Washington, D.C.: GPO, 1947, No. 5. Available in public technical libraries.

116. Parish, J.A., & Drucker, A.J. Personnel Research for Officer Candidate School, USA TAGO Personnel Research Branch Technical Research Report, 1957, No. 117, 22pp. Available in public technical libraries.

117. H. Roy, J. Brueckel, and A.J. Drucker, "Selection of Army and Air Force Reserve Training Corps Students," USA Personnel Research Branch Notes 28, 9 (1954). Available in public technical libraries.

118. Standard Oil Company of New Jersey. Social Science Research Reports: Selection and Placement, Vol. II. New York: Author, 1962.

119. W.D. Buel, "Voluntary Female Clerical Turnover: The Concurrent and Predictive Validity of A Weighted Application Blank," Journal of Applied Psychology 48, 180-182 (1964). Available in public technical libraries.

120. M. Baehr, J.E. Furcon, and E.C. Friend, Psychological Assessment of Patrolman Qualifications In Relation to Field Performance. Washington, D.C.: Superintendent of Documents, US GPO, 1968.

121. R. Lee and J.M. Booth, "A Utility Analysis of A Weighted Application Blank Designed to Predict Turnover For Clerical Employees," Journal of Applied Psychology 59, 516-518 (1974). Available in public technical libraries.

122. J.A. Plag and J.M. Goffman, "The Armed Forces Qualification Test: Its Validity in Predicting Military Effectiveness for Naval Enlistees," Personnel Psychology 20, 323-329 (1967). Available in public technical libraries.

123. W.J. Smith, L.E. Albright, J.R. Glennon, and W.A. Owens, "The Prediction of Research Competence and Creativity from Personal History," Journal of Applied Psychology 45, 59-62 (1961). Available in public technical libraries.

124. W.F. Glueck, Personnel: A Diagnostic Approach. Business Publications, Inc. Dallas, 1974.

125. L.A. Pace and L.D. Schoenfeldt, "Legal Concerns In the Use of Weighted Applications," Personnel Psychology 30, 159-166 (1977). Available in public technical libraries.

126. R.M. Liebert and M.D. Spiegler, Personality: An Introduction to Theory and Research, The Dorsey Press. Homewood, Illinois, 1970.

127. M.M. Linehan, "Issues in Behavioral Interviewing," in Behavioral Assessment: New Directions in Clinical Psychology, J.D. Cone and R.P. Hawkins, Eds. (Brunner/Mazel, Inc., New York, 1977).

128. A. Crosby, "The Psychological Examination in Police Selection," Journal of Police Science and Administration 7, 215-229 (1979). Available in public technical libraries.

129. C.J.G. Perry, "Psychiatric Selection of Candidates for Space Missions," Journal of the American Medical Association 194, 841-844 (1965). Available in public technical libraries.

130. W.F. Cascio and L.J. Real, "The Civil Service Exam Has Been Passed: Now What?" in Police Selection and Evaluation, C.D. Spielberger, Ed. (Praeger Publishers, New York, 1979).

131. R. Wagner, "The Employment Interview: A Critical Review," Personnel Psychology 2, 17-46 (1949). Available in public technical libraries.

132. E.C. Mayfield, "The Selection: A Re-Evaluation of Published Research," Personnel Psychology 17, 239-260 (1964). Available in public technical libraries.

133. L. Ulrich and D. Trumbo, "The Selection Interview Since 1949," Psychological Bulletin 63, 100-116 (1965). Available in public technical libraries.

134. E.D. Webster, Ed. Decision Making in the Employment Interview. Eagle, Montreal, 1964.

135. M.H. Palacios, L.A. Newberry and R.R. Bootzin, "The Predictive Validity of the Interview," Journal of Applied Psychology 50, 67-72 (1966). Available in public technical libraries.

136. J.B. Miner, "Executive and Personnel Interviews As Predictors of Consulting Success," Personnel Psychology 23, 521-538 (1970). Available in public libraries.

137. P. Stringer and M. Tyson, "University Selection Interviewer's Ratings Related To Interviewee Self-Image," Occupational Psychology 42, 49-60 (1968). Available in public technical libraries.

138. F.L. Landy, "The Validity of the Interview in Police Officer Selection," Journal of Applied Psychology 61, 193-198 (1976). Available in public technical libraries.

139. M.D. Hakel, "Similarity of Post-Interview Trait Rating Intercorrelations as A Contributor to Inter-Rater Agreement In A Structured Employment Interview," Journal of Applied Psychology 55, 443-448 (1971). Available in public technical libraries.

140. R.M. Guion, "Recruiting, Selection, and Job Replacement," in Handbook of Industrial and Organizational Psychology, M.D. Dunette, Ed. (Rand McNally College Publishing Company, Chicago, 1976).

141. F.J. Landy and D.A. Trumbo, Psychology of Work Behaviors. The Dorsey Press, Homewood, Illinois, 1976.

142. D.P. Schwab and H.G. Heneman, "Relationship Between Interview Structure and Inter-Interviewer Reliability In An Employment Situation," Journal of Applied Psychology 53, 214-217 (1969). Available in public technical libraries.

143. W.D. Scott, R.C. Clothier, and W.R. Spriegel, Personnel Management: Principles, Practices and Point of View. McGraw-Hill, New York, 1949.

144. J. Mosel and H. Goheen, "Validity of the Employment Recommendation Questionnaire In Personnel Selection," Personnel Psychology 11, 481-490 (1958). Available in public technical libraries.

145. K.R. Pelletier, Mind as Healer, Mind as Slayer, Dell Publishing Company. New York, 1977.

146. T.H. Holmes and R.H. Rahe, "The Social Readjustment Rating Scale," Journal of Psychosomatic Research 11, 213-218 (1967). Available in public technical libraries.

147. E.S. Paykel, B.A. Prusoff, and E.H. Uhlenhuth, "Scaling of Life Events," Archives of General Psychiatry 25, 340-347 (1971). Available in public technical libraries.

148. R.H. Rahe, "Life-change Measurement as a predictor of Illness," Proceedings of the Royal Society of Medicine 61, 1124-1126 (1973). Available in public technical libraries.

149. T.H. Holmes and M. Masuda, "Life Change and Illness Susceptibility, Separation and Depression," A.A.A.S., 161-186 (1973). Available in public technical libraries.

150. A.L. Komaroff, M. Masuda, and T.H. Holmes, "The Social Readjustment Rating Scale: A Comparative Study of Negro, Mexican and White Americans," Journal of Psychosomatic Research 12, 121-128 (1968). Available in public technical libraries.

151. M. Masuda and T.H. Holmes, "The Social Readjustment Rating Scale: A Cross-cultural Study of Japanese and Americans," Journal of Psychosomatic Research 11, 227-237 (1967). Available in public technical libraries.

152. N.K. Walker, "The Application of Dual Task (ZITA/ADT) Stress Testing to the Selection of Personnel for Nuclear Power Stations," Report No. 94, submitted to the USN RC, 1980. Available for purchase from National Technical Information Service, Springfield, VA 22161.

153. "Stress Device Aides Medicine," Washington Post, June 13, 1974. Available in public libraries.

154. N.K. Walker and M.M. Walker, "Dual-task (ZITA/ADT) Performance of the Fighter Pilot Population (based on 84 pilots in lead-in training)," Report No. 92, DARPA order 1001/983. Norman K. Walker Associates, Inc. Rockville, Md., 1980. Available for purchase from National Technical Information Service, Springfield, VA 22161.

7. FINAL STANDARDS FOR COMPARISON OF PSYCHOLOGICAL ASSESSMENT TECHNIQUES FOR NUCLEAR FACILITY PERSONNEL SELECTION

7.1 Introduction

Prior to the recommendation of specific techniques/instruments for use in nuclear facility personnel selection procedures, it is necessary to delineate acceptable standards for instrument adherence. These standards represent the criteria which should be met in order to provide legal, ethical, statistical and professional justification for the procedure(s) of choice, regardless of the specific nuclear facility position in question.

Due to the diversity of factors relevant to the specification of particular selection instruments, the standards established were derived from a variety of sources, including the following:

- Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, Department of Justice. Uniform Guidelines on Employee Selection Procedures. Federal Register, 1978, 43, (166), 38289-38315.

- American Psychological Association, Standards for Educational and Psychological Tests (Rev. ed.). Washington, D.C., APA, 1974.

- American Psychological Association, Division of Industrial-Organizational Psychology. Principles for the Validation and Use of Personnel Selection Procedures. (Second ed.). Berkeley, CA, APA, 1980.

- American Psychological Association, Ethical Standards of Psychologists (Rev ed.). Washington, D.C., APA, 1977.

- Considerations applied to selection instruments reviewed in the original Literature Document submitted under NRC Contract Number NRC-01-79-002 (See Section 5.3 of this report).

- Considerations, comments, suggestions, and issue resolutions contributed by members of an interdisciplinary panel of experts, each associated with the nuclear industry in a particular capacity. This Criteria Development Panel Meeting convened in Orlando, Florida, on 8/22/80.

The standards outlined fall into two major categories: 1) those standards which apply to individual instruments where evaluation of available instruments may reveal dif-

fering degrees of adherence to each standard, and 2) those standards which are independent of individual instrument characteristics and which should be satisfied by the total selection process.

These standards are representative of ideal instrument compliance. Our review of the various instruments presently available did not reveal any measures which currently demonstrate adequate adherence to all of the standards set forth. This finding has implications for the need for further research designed to evaluate relative degrees of instrument and procedural compliance with the established standards (see Chapter 10 for a discussion of needed research).

7.2 Standards Applicable to Individual Instruments

The standards which follow are based on the preceding introductory material, and are divided into major categories of criteria, each having a number of separate components:

7.2.1 Reliability

As stated in the American Psychological Association's Standards for Educational and Psychological Tests "Reliability refers to the degree to which the results of testing are attributable to systematic sources of variance." (p. 48). In more succinct terms, reliability refers to the consistency of a test. Estimation of reliability coefficients is traditionally accomplished by correlating data from a minimum of two sets of similar measurements. There are a variety of methods for deriving these correlations, each appropriate for a given application. For the purposes of nuclear facility personnel selection instruments, the recommended standards are as follows:

- Since hiring decisions are based on recommendations derived from various data sources that are typically combined by the professional who makes the final selection recommendation, inter-rater reliability

should be the key index in establishing instrument/procedure reliability.* Inter-rater reliability is a measure of the extent to which two or more examiners are in agreement in their evaluation of the same test specimen. Thus, if there is a high level of agreement among different raters' measures of an applicant's emotional stability, there is said to be a high inter-rater reliability.

- With respect to inter-rater reliability, two reliabilities should be computed:

  (1) Inter-rater reliability for all decisions, i.e., both correct rejects and acceptances. In this case, the desirable inter-rater reliability would be .80 to .90.

  (2) Inter-rater reliability focusing only on those applicants rejected. The desirable inter-rater reliability for this situation would be .50, as the base rate for correctly rejecting an applicant at random would be quite low.

- If more than one form of the instrument is available, studies should be conducted to demonstrate that multiple forms measure the same factors, have equal levels of difficulty, require equal administration times, etc. Such investigations yield measures of equivalent forms reliability.

---

* It is recognized that many of the personality tests currently available, including those reviewed in Chapter 6, have not used the inter-rater approach to measuring the reliability of these instruments. This is largely due to the fact that clinicians typically combine the results of clinical interviews and personality test results in reaching conclusions. Thus, inter-rater reliability data relating specifically to personality tests have historically been unavailable. It is therefore recommended that research using the inter-rater reliability model be conducted by employing "blind" interpreters with no information about the applicants other than those personality test profiles they are given. Comparisons of these interpretations would yield inter-rater reliability data for personality tests that have not been confounded by other variables.

- Reliability studies should include complete descriptions of the sample groups and statistical procedures used to determine reliability coefficients. Such information would be necessary in helping a potential user of the instrument to determine whether the results obtained from the sample study would be applicable to the population to be measured.

- Reliability studies of specific instruments should include information regarding the standard error of measurement of the instrument. Whereas the reliability coefficient provides an estimate of an instrument's consistency, the standard error of measurement, a derivation of the reliability, provides an estimate of how accurate an individual's test score is.

7.2.2 Validity

Validity may be defined as the degree to which an instrument does, in fact, measure what it purports to measure. In the area of personnel selection, the notion of validity refers to the relationship between a selection instrument and the job performance it is intended to measure.

There are several approaches to the measurement of validity, which are briefly described below:

- Criterion-Related Validity*: This procedure demonstrates validity by establishing a relationship between an individual's scores on a selection instrument and some measure of his/her performance or behavior (in this case, job performance). The selection instrument is referred to as the predictor, and the measure of job performance is known as the criterion.

- Content Validity: With this approach, a selection instrument is validated by evidence demonstrating that the procedure representatively samples important elements of the job.

* Criterion-related validity is often used interchangeably with the term criterion-oriented validity.

- Construct Validity: This validation method is based on the identification of a psychological trait or construct, which is presumed to be related to successful job performance. A selection procedure designed to measure the presence and degree of this trait in the individual is then devised. Validation of this strategy is accomplished by comparison of the selection instrument scores to other measurement devices presumed to measure the same trait.

For the purposes of nuclear facility personnel selection, the following standards are recommended for the validation of psychological assessment instruments:

- Wherever possible, validation strategies should utilize the predictive criterion-related validity model.

- The criteria should be job-related (based on reviews of job analyses) and should be accurately and completely described.

- If the criterion-related validity approach is not feasible, the user may conduct studies based on content validity or construct validity approaches.

- In conducting content validity studies, the researcher must accurately and completely define the performance domain. This domain must be a representative sample of required job behaviors, as based on the job analysis.

- When the construct validity approach is used, each construct should be named and defined in terms of a particular theoretical interpretation. The accompanying job analysis should include an identification of constructs believed to underlie the successful performance of critical work behaviors.

- The sample groups used in validity studies should be consistent with the application to nuclear facility personnel selection.

- Validity study reports should include complete descriptions of the experimental and statistical procedures employed.

- Validity study reports should present any evidence of bias due to ethnic, sex, or other sub-sample differences.

- Where more than one instrument is to be used in a selection procedure, validity data should be obtained for the combination procedure as a whole, as well as for the individual components.

7.2.3 Compliance with Legal Issues, Labor Relations, and Uniform Guidelines on Employee Selection Procedures

- Wherever possible, criterion-related validity should be established for selection procedures used, in order to assure compliance with EEOC Guidelines.

- The instruments should involve procedures which are as work-related as possible.

7.2.4 Personal Effects on Applicants

- Use of the instrument should not constitute an invasion of privacy.

- There should be no severe personality or physical reactions resulting from the administration of the instrument.

- The use of the instrument should not be so unobtrusive that the evaluatee is unaware that he/she is being evaluated.

7.2.5 Susceptibility to Faking

- Instruments used in selection procedures should employ some type of system designed to detect and/or safeguard against applicant distortion of results during the administration phase. This would primarily be the case with instruments where research indicates susceptibility to faking (see Chapter 11 for a discussion of procedures to be followed if faking is detected during the course of a selection procedure).

7.2.6 Reevaluation Considerations

- Scores on the instrument or on another form of the instrument should not, in any fashion, be influenced by results of the first administration.

- Research findings attesting to the above point should be available.

- Acceptable alternative forms should be available for instruments where reevaluation results would be contaminated by initial evaluation. In such cases, there should be manuals or research reports

accompanying the instruments which provide data on means, variances, characteristics of items included in the forms, and coefficients of correlation among the various forms.

7.3 Standards for Adherence with Respect to the Total Selection Procedure

The standards which follow are those which should be mandatory for the entire selection process.

7.3.1 Administrator/Examiner: Training and Qualifications

- The administrator should have a general knowledge of measurement principles and the limitations of test interpretation.
- The administrator should be acquainted with the literature relevant to the instrument he/she is using, as well as being thoroughly trained in its administration and interpretation.
- In order to assure administrator competence, investigations of intra-rater reliability should be conducted for the instruments of choice. Such studies should focus on individual examiner's blind scoring/rating/interpretation of a given applicant's instrument results at one point in time, compared with the same examiner's evaluation of the same applicant's original instrument results at a later date.

7.3.2 Confidentiality

- Prior to participating in any phase of the selection procedure, applicants should sign informed consent forms. These forms should clearly state what type of information is to be provided to what specific individuals, and the specific purposes of this information.
- The responsible professional should interpret test results in a format understandable by laypersons before releasing these data. The information provided should be clearly related to the specific job in question.
- The professional should refrain from releasing any unnecessary information that is potentially damaging to the applicant.
- The professional should develop a system for removing from data files selection instrument results that have become obsolete due to the passage of time.

7.3.3 Subsequent Evaluations of Applicants

- In accordance with Uniform Guideline provisions, procedures for reevaluation of those applicants screened out should be established.

- The responsible professional should establish appropriate time intervals between initial evaluation and reevaluation for applicants screened out. Determination of these time intervals should be based on initial screening results and the life of the data.

7.3.4 Other Considerations in Regard to Applicants

- The applicant should be informed of all instruments to be used in the selection process prior to any instrument administration, and should agree to this process by signing an informed consent form.

- Results of selection procedures should be viewed as predictions of performance in specific contexts and situations. They should not be treated as measures of absolute traits of an individual, nor should they be seen as permanently enduring characteristics which are present in all settings, and at all times.

- Feedback of selection procedure results to applicants is critical. The nature of this feedback, particularly for an applicant screened out, is also of vital importance. Feedback should be phrased in terms of findings which are specifically relevant to job situations, rather than indicative of general degrees of emotional instability.

- If an applicant is screened out, the professional who makes this decision should provide the individual with remedial action alternatives for resolving the problems disclosed.

7.4 Additional Considerations

The Instrument Review section of this report (Chapter 6) included the following three additional instrument considerations that are not included in the Standards Sections (7.2 and 7.3) of this report:

(1) Cost of Development and Administration
(2) Equipment/Materials Needed for Administration
(3) Administration: Time and Difficulty

our decision to exclude these factors as formal standards for instrument consideration was based on relevant discussions held at the Criteria Development Panel Meeting, as well as our attention to the critical importance of choosing the most appropriate instruments for selection procedures. It was the joint consensus of ADI personnel and the panel meeting members that it would be impossible to establish fixed cost, equipment, time and difficulty standards, since these elements are not critical to the determination of which instruments would most accurately detect and/or predict emotional instability within the confines of psychometric, professional, legal, and ethical requirements. Thus, the choice of the most appropriate instruments should be made on the basis of adherence to the standard criteria previously presented, after which these additional elements can be considered on the basis of the instruments of choice.

7.5 Matrix 1: Measurement Instruments by Standards for Comparison of Psychological Assessment Techniques

7.5.1 Introduction

The purpose of the matrix included in this section is to provide a concise means of evaluating the measurement instruments reviewed in Chapter 6 in terms of their relative degrees of adherence to the individual instrument standards outlined in Section 7.2 of this chapter. Standards focusing on the total selection process (Section 7.3) are not included in this matrix.

7.5.2 Methodology

Each instrument reviewed in Chapter 6 is assigned a rating indicating degree of adherence to each standard included in the matrix. The numerical ratings on the matrix were derived by the following procedure:

(1) A numerical rating system was devised with ratings of 1, 2, and 3. A rating of 1 implies minimal instrument adherence to the standards; 2 implies moderate adherence; and 3 implies a high degree of adherence to the standards developed.

(2) Each standard was individually examined in terms of how the numerical ratings were applicable to that particular standard (Section 7.5.3 outlines this interpretation of the ratings with respect to each standard).

(3) On the basis of the instrument reviews presented in Chapter 6, ratings for each standard were assigned to each of the instruments on the basis of the instrument's adherence to the provisions of each standard. These ratings represent the collective judgment of the ADI staff members who composed this report.

7.5.3 Interpretation of Ratings for Each Standard

The following information outlines the meanings of the ratings for each standard. The reader should become thoroughly familiar with this interpretive information prior to examining the matrix. In addition to ratings of 1, 2, and 3, some instruments received non-numerical ratings for some standards. These ratings include:

(1) NA: this standard is not applicable to this instrument.

(2) A blank cell indicates that inconclusive evidence was available to this review for rating on a particular standard.

Interpretation of the numerical ratings for each standard is as follows:

(A) Reliability*

3 = High mean reliability coefficient for this instrument

2 = Moderate mean reliability coefficient for this instrument

1 = Low mean reliability coefficient for this measurement

* On the matrix, reliability is broken down into categories of Test-Retest, Equivalent Forms, Internal Consistency, and Inter-Rater Reliability due to variations in applicability of the various types of reliability to different instruments.

(B) Validity*

3 = High mean validity for this instrument

2 = Moderate mean validity for this instrument

1 = Low mean validity for this instrument

(C) Compliance with Legal Issues, Labor Relations and Uniform Guidelines on Employee Selection Procedures

3 = Well-established validity for this instrument; evidence demonstrating that the procedure is work-related; no evidence of probable adverse impact.

2 = Moderate evidence for validity of this instrument; procedure appears somewhat work-related; may have potential for adverse impact.

1 = No well-established validity of any type (based on available data pertaining to instrument); procedure does not appear work-related; high probability of resulting in adverse impact.

(D) Personal Effects on Applicants

3 = Use of the instrument does not constitute an invasion of privacy; does not induce any severe psychological or physical reactions, and is not so unobstrusive that the applicant is unaware of being tested.

* On the matrix, validity is broken down into categories of criterion-related, content, and construct validity, due to variations in applicability of the various types of validity to different instruments. The ratings presented for criterion-related validity refer only to the validity of the instrument with respect to its measuring behavioral on-the-job indices of emotional instability.

2 = Use of the instrument could possibly constitute an invasion of privacy, may induce some degree of psychological or physical reactions; and may be so unobstrusive that the applicant is unaware of being tested.

1 = Use of the instrument probably constitutes an invasion of privacy; is likely to induce significant psychological or physical reactions in applicants; and may be so unobstrusive that applicant is unaware of being tested.

(E) Susceptibility to Faking

3 = The instrument has a well-designed system which is effective in detecting or safeguarding against applicant distortion of results.

2 = The instrument has some type of mechanism designed to detect applicant distortion of results, but may not always be very efficient in this regard.

1 = The instrument has no system designed to detect applicant distortion of results.

(F) Reevaluation Considerations

3 = The instrument is designed in such a way that results on subsequent administrations of the instrument (or another form of it) are not influenced by results from the first administration, and there are research findings attesting to this point; acceptable alternative forms of the instrument with well-documented research support of equivalence are available.

2 = The instrument is designed in such a way that initial results may have a minimal effect on subsequent evaluations of the same instrument or another form of it; alternative forms are available, but supportive research is limited.

1 = The instrument is designed in such a way that initial results and administration procedures are highly likely to influence results of subsequent administration; alternative forms of the instruments are not available.

Ma x 1

Measurement Instruments by Standards
for Comparison of Psychological Assessment Techniques

MEASUREMENT INSTRUMENTS	TEST-RETEST RELIABILITY	EQUIVALENT FORMS RELIABILITY	INTERNAL CONSISTENCY RELIABILITY	INTER-RATER RELIABILITY	CRITERION-RELATED VALIDITY*	CONTENT VALIDITY
MMPI (Minnesota Multiphasic Personality Inventory)	3	( )	1	( )	( )	NA
CPI (California Psychological Inventory)	2	( )	( )	( )	( )	NA
16PF (Sixteen Personality Factor Questionnaire)	1	1	( )	( )	2	NA
GPP-I (Gordon Personal Profile Inventory)	2	( )	3	( )	( )	NA
EPQ (Eysenck Personality Questionnaire)	3	( )	3	( )	( )	NA
Biofeedback	( )	( )	( )	( )	( )	NA
Polygraph	1	( )	( )	2	( )	NA
Telemetry	( )	( )	( )	( )	( )	NA
Voice Stress Analyzer	( )	( )	1	2	( )	NA
Structured Pre-Employment Interview	( )	( )	( )	2	( )	( )
Unstructured Pre-Employment Interview	( )	( )	( )	2	( )	( )
Stress Pre-Employment Interview	( )	( )	( )	2	( )	( )
Letters of Reference	( )	( )	( )	( )	( )	( )
Background Check	( )	( )	( )	( )	( )	( )
Field Investigation	( )	( )	( )	( )	( )	( )
Situational Simulations	[illegible]	( )	2	3	( )	3
Weighted Application Blanks	3	( )	2	( )	( )	( )
Clinical Interview	( )	( )	( )	1	( )	NA
Life Change Scales	[illegible]	( )	( )	( )	( )	NA
ZITA/ADT	( )	( )	2	( )	( )	( )

KEY

3 = High degree of adherence to standard
2 = Moderate degree of adherence to standard
1 = Minimal degree of adherence to standard
NA = Not applicable to this instrument
(A blank cell) = Inconclusive evidence available for rating on this standard.

Matrix 1 (continued)

MEASUREMENT INSTRUMENTS	CONSTRUCT VALIDITY	COMPLIANCE WITH LEGAL ISSUES LABOR RELATIONS, AND UNIFORM GUIDELINES ON EMPLOYEE SELECTION PROCEDURES	NEGATIVE EFFECTS ON APPLICANTS	SUSCEPTIBILITY TO FAKING	REEVALUATION CONSIDERATIONS
MMPI (Minnesota Multiphasic Personality Inventory)	3	2	3	3	2
CPI (California Psychological Inventory)	2	2	3	3	1
16PF (Sixteen Personality Factor Questionnaire)	2	2	3	3	3
GPP-I (Gordon Personal Profile Inventory)	2	2	3	1	2
EPQ (Eysenck Personality Questionnaire)	2	2	3	2	1
Biofeedback	( )	( )	1	2	2
Polygraph	1	1	1	2	1
Telemetry	( )	( )	1	3	( )
Voice Stress Analyzer	1	1	1	3	1
Structured Pre-Employment Interview	2	3	3	1	2
Unstructured Pre-Employment Interview	2	3	3	1	2
Stress Pre-Employment Interview	2	3	3	1	1
Letters of Reference	( )	3	3	1	NA
Background Check	( )	2	2	2	NA
Field Investigation	( )	1	1	2	NA
Situational Simulations	2	3	3	3	2
Weighted Application Blanks	3	2	3	2	2
Clinical Interview	( )	1	3	2	2
Life Change Scales	( )	( )	3	2	2
ZITA/ADT	( )	( )	3	3	1

KEY

3 = High degree of adherence to standard
2 = Moderate degree of adherence to standard
1 = Minimal degree of adherence to standard
NA = Not applicable to this instrument
(A blank cell) = Inconclusive evidence available for rating on this standard.

8. RECOMMENDED SELECTION PROCEDURE FOR NUCLEAR SECURITY PERSONNEL

8.1 Introduction

The material contained in this section is based upon the information compiled throughout all phases of the project. The recommendations included here represent the best judgment of ADI staff members who participated in the project, based on available data, and inputs from experts in the field.

As reflected in Chapter 6, no one measurement instrument currently meets all the standards for instrument consideration (as outlined in Chapter 7). The instruments recommended here are those which best met the "critical" standard criteria. After listing the components of our recommended procedure, each separate instrument/technique will be discussed to explain the bases for our decisions. While the discussion focuses on the critical standard criteria, other considerations bearing on each instrument are also addressed.

The selection procedure recommended by ADI includes the following:

(1) Standardized clinical interviews

(2) Administration of the Minnesota Multiphasic Personality Inventory (MMPI) and the Sixteen Personality Factor Questionnaire (16PF).

(3) Situational simulations appropriate for each of the two security positions addressed, i.e., security guard and supervisor.

A professional (see Chapter 11 for Professional Standards and Qualifications) should combine the results of these diverse procedures, and provide, in lay terms, the results of the evaluation process to the hiring official(s) within the organization. The results should be documented on a standard form which lists the instruments administered and presents the results in job-related, behavioral terms (see Appendix E for sample form).

8.2 Discussion of Specific Instruments Recommended

8.2.1 Clinical Interviews

Although more research pertaining to the reliability and validity of the clinical interview is needed, it has traditionally been one of the chief ingredients in the overall evaluation process in the measurement of emotional stability.

An effective clinical interview, characterized by well-positioned probing, can detect indices of emotional instability, which otherwise might not be revealed. The inclusion of the clinical interview in the evaluation process is based upon a number of considerations.

- It is the most widely used method of assessing personality.
- It allows for further in-depth probing of behavioral tendencies initially detected through other means.
- It allows for clarification of inconsistencies in data revealed through other means.
- Although available research findings are not particularly supportive of its inter-rater reliability, its approach lends itself to this evaluation strategy.
- Although available research findings are not particularly supportive of its inter-rater reliability, the structured format offers promise for acceptable inter-rater reliability.
- It lends itself to a construct validation strategy.
- Inclusion of certain types of questions in the interview can detect faking on the part of the applicant.
- Carrying out re-evaluations on an applicant are appropriate, given that a second professional is involved in the re-evaluation phase.
- In terms of personal effects on applicants, given the use of proper interviewing techniques and proper treatment of the data, there should be minimal potential for detrimental effects to applicants.
- The clinical interview is the point of integration of all data obtained during the entire evaluation process.

A structured clinical interview methodology is recommended. The format for the interview should be well-defined, and should contain the following areas of questioning:

- Demographic traits
- Marital status, family composition, and any problems associated therewith
- Any unusual aspects of developmental history.
- Any recurring vivid or traumatic childhood memories
- Any history of substance abuse/dependency
- Any recent changes in mood, concentration abilities, appetite, sleep patterns, general health, etc.
- Any unusual fears or areas of concern
- Any strong feelings regarding the use of nuclear power, either pro or con
- Any specific concerns or doubts regarding job for which applicant is applying
- Any inconsistencies or problem traits and behaviors revealed by other selection instruments

8.2.2 Personality Tests: the MMPI and the 16PF

Personality tests are a relatively low-cost, easily administered technique used by professionals in combination with information obtained through clinical interviews. The particular tests recommended and their respective justifications are as follows:

(1) MMPI: The MMPI is the most well-established and widely used personality inventory available. Its inclusion in the evaluation process is based upon a number of considerations.

- It is a comprehensive test designed to measure indices of all the major categories of aberrant behavior.
- It has been the subject of more research studies than any other personality inventory.

- There are some acceptable reliability data, although most of these focus on test-retest reliability.

- There is good evidence of construct validity.

- There have been some criterion-related validity studies with occupational groups, although the correlations have not been uniformly supportive.

- Normative data exist on sub-groups in the population.

- Scales are included for detecting faking as well as other response sets which might invalidate test results.

- Its use as a selection device has a historical basis, as a number of organizations have used this instrument in their personnel selection procedures.

- In terms of personal effects on applicants, assuming proper and confidential treatment of the data, there is minimal potential for detrimental effects to applicants.

- As it is familiar to most professionals, less special training would be needed to incorporate this instrument into a standard selection system.

(2) 16PF: The 16PF is a comprehensive personality inventory designed to provide an individual's scores on what Cattell felt were the sixteen basic or "source" traits. The test provides scores along a continuum between the two poles of each source trait. Its inclusion in the evaluation process is based upon a number of considerations.

- It is a very widely used personality inventory, particularly with occupational groupings.

- There are some acceptable reliability data, although most of these focus on equivalent forms and test-retest aspects.

- There is some support for criterion-oriented validity, although more long-term research would be appropriate.

- There has been some extensive construct validity research, focusing on "stress" present in jobs similar to that found in the nuclear industry context.

- The development of the test was based on sound research utilizing test construction principles.

- The test was constructed primarily to measure interpersonal relations dimensions of the normal adult personality, which would be expected to be exhibited by the majority of the job applicant pool.

- Scales are included for detecting faking as well as other factors which might invalidate test results.

- It has five forms, which satisfy re-evaluation considerations.

- In terms of personal effects to applicants, assuming proper treatment of the data, the potential for detrimental effects to applicants appears minimal.

8.2.3 Situational Simulations

While not located in vital or protected areas, security positions are highly stressful in nature. Thus, although personality inventories and clinical interviews are necessary tools for detecting aberrant behavioral tendencies, there is, in addition, the need to measure one's ability to react to on-the-job factors associated with tolerance of boredom, crisis elements, etc. Elements such as these can be best measured through the situational simulation approach. Its inclusion in the evaluation process is based upon a number of considerations.

- It is a well-accepted technique used by many organizations within their personnel selection procedures.

- There is support in terms of its inter-rater reliability.

- It has a sound content validity basis, given that instruments are properly developed.

- There is evidence of criterion-oriented validity, although more relevant, long-term research would be appropriate.
- There is some evidence for its construct validity.
- There are available some research findings pertaining to sub-group validity.
- In terms of legal considerations, it has been tested and upheld in court.
- Research indicates that it is difficult to fake scores obtained through this approach.
- It has been used extensively in organizations which bear resemblance to the nuclear power industry.
- If the simulations are properly developed, it will measure on-the-job behaviors, e.g., tolerance of boredom, not easily measured through other means.
- In terms of personal effects to applicants, given appropriate design of the simulations and proper treatment of the data, the potential for detrimental effects to applicants appears minimal.

8.2.3.1. Nature of the Simulations

While it is not our purpose to specify the precise nature of potential simulation devices, certain factors should be addressed.

(1) Content*: Consideration should be given to such elements as:

- Requiring evaluatees to perform monotonous tasks for an unspecified period of time.

* The exact content of the simulations would vary from position to position and thus, at this time, in-depth discussion of possible simulations would be unproductive. Elements listed here would be particularly appropriate for a security position.

- After performing monotonous tasks, suddenly shifting to a more active involvement.

- Responding to a diversion, different from the normal (hypothetically speaking) work responsibilities.

- Responding to a simulated "crisis" situation where reaction time, and other factors would be evaluated.

Specifically with respect to simulating boredom, while it would be difficult to simulate precisely the cumulative effects of boredom, certainly a reasonable facsimile could be created.

(2) Evaluators:

- Evaluators could be either persons outside the organization or full-time employees of the organization. In either case, quality evaluator training would be a necessity. Training would be required in the observation, classification, and evaluation of behavior, through the use of simulation technology.

(3) Time Parameters

- Four (4) hours of evaluatee time would be necessary

- Approximately six (6) hours of evaluator time (per evaluatee) would be necessary

(4) Features of the Simulation

- Depending upon the given position (i.e., Security Guard or Supervisor) for which the evaluatee is applying, the degree to which the conditions in which the simulation would be embedded approximate those of the job, would vary. Elements to consider would be the importance of preserving condition fidelity with respect to the job and cost factors. For some positions, it would be appropriate to isolate a segment of the facility in which to conduct the evaluation. Such considerations would also involve the determination of necessary equipment.

(5) Relationship to other on-going use of simulation technology:

- It is recognized that simulations are currently used for other purposes within the nuclear industry, e.g., training. While it is appreciated that redundancy of effort and cost are to be avoided, it is also stressed that if the types of simulations presently being used are not appropriate for a given job, or not appropriate for measuring emotional stability within a selection context, then innovative technology would be necessary.

- With respect to and consistent with the above point, consideration should be given to the use of simulations for some positions, during the probationary period of employment. In this case, the determination of permanent employment would be made after the evaluation of emotional stability during training. Factors to be considered in choosing this strategy would be cost aspects of training and temporarily employing some people who would eventually be rejected on the basis of emotional instability, versus cost savings in not having to put all applicants through the simulation process.

8.3 Measure Redundancy

While it has been recommended that the selection system consist of four components, the MMPI, the 16PF, situational simulations and the clinical interview, it is also recognized that some interested parties might be concerned with measure redundancy present in the system. It is our collective opinion that each of these instruments contributes in a unique fashion to the prediction of emotional instability, with some slight redundancy present with respect to the MMPI and 16PF. If the NRC wishes to reduce the number of components, the only potential area for instrument elimination would be with respect to the MMPI or 16PF. Careful consideration would be advisable, however, in the decision to eliminate either the MMPI or the 16PF from the selection procedure. Despite some possible measure redundancy, these instruments complement each other in several ways and have both been recommended for the following reasons:

(1) The MMPI was originally normed on a psychiatric population, thus it is primarily designed to detect "abnormal" behavioral patterns including indices of moderate to extreme emotional instability. This is reflected in the psychiatric terminology used in naming the scales, which

are primarily representative of categories of mental and emotional pathology.* The "normal" limits of behavior are usually not subject to any in-depth interpretation. Thus, in a nuclear facility selection procedure, the MMPI would be expected to detect those individuals in the moderate to extreme range of emotional instability, including those who might otherwise be able to conceal their unstable tendencies if they were not subjected to evaluation.

In contrast, the 16PF was developed for use primarily with the "normal" adult population, and was intended more for applied uses as opposed to the MMPI's main objective being the detection and prediction of aberrant traits and behaviors from a clinical perspective. The 16PF was constructed to provide an overall picture of an individual's personality in terms of major traits and behavioral tendencies. Due to its focus on the normal range of behavior, it appears to be capable of detecting subtle indices of emotional instability which might hinder on-the-job performance. The 16PF includes scales which are more oriented toward interpersonal interaction styles than those on the MMPI. Knowledge of how an individual interacts with others is a vital source of information in a personnel selection procedure focusing on behavioral manifestations of emotional instability.

(2) The MMPI is the most widely used personality inventory due to its comprehensive nature to detect an extensive range of inappropriate behavioral tendencies. Because of this extensive range of evaluation, the MMPI is relatively lengthy, and usually takes from forty-five minutes to an hour for administration. In comparison, the scales on the 16PF are shorter, thus administration time is only approximately fifteen minutes. The time factors associated with these two instruments would demand less of the applicant's and administrator's time than some other combinations of tests, while at the same time providing complementary information.

*The psychiatric terminology employed does not detract from applied uses of the MMPI due to the availability of interpretive data designed to predict overt behaviors.

8.4 Matrix 2: Measurement Instruments by Indices of Emotiona Instability

In order to more clearly and concisely depict which indices of emotional instability may be detected or predicted by the instruments recommended, the following matrix is provided. The categories of emotional instability indices used here are those previously outlined in Section 3.2.1. An "X" within a cell indicates that the measurement instrument listed in the vertical column can detect or predict the category of emotional instability indices listed in the horizontal column. This matrix is presented on the following page.

MATRIX 2

Measurement Instruments by Indices of
Emotional Instability

Categories of Emotional Instability Indices	Measurement Instruments: Clinical Interviews	MMPI	16PF	Situational Simulations
Immediate or Short-Term Reactions to Crisis Situations				X
Reactions to Long-Term Effects of Accumulated Stress				X
Hostility Toward Authority	X	X		
Illegal and Antisocial Behaviors		X	X	
Irresponsibility	X	X	X	X
Dependent Behavioral Patterns			X	
Interpersonal Skill Deficiencies	X	X	X	X
Deficiencies in Vigilance			X	X
Emotional and Thought Disturbances	X	X		

9. APPLICABILITY OF RECOMMENDED SELECTION PROCEDURES TO OTHER NUCLEAR FACILITY POSITIONS

9.1 Introduction

9.1.1 Objectives: The main objective of the extension to our original contract was to determine the extent to which the selection procedures recommended for nuclear security personnel were generalizable to other positions within nuclear facilities. More specifically, the objectives were:

(1) The identification of critical job characteristics

   (a) relating to emotional instability, and
   (b) pertaining to other occupations within nuclear facilities.

(2) The determination of whether the standards previously developed (see Chapter 7), and the selection procedures recommended for security personnel (see Chapter 8) in accordance with these standards are equally applicable to all positions within nuclear facilities.

9.1.2 Informational Sources

In conducting this phase of the project, there was a reliance upon several sources of information. These were:

- Documents and materials previously referred to (see Chapter 2 for a listing of these sources).
- Job description information on typical positions found within a nuclear facility which was provided by nuclear industry organizations.
- The classifications of jobs enumerated in NUREG/CR-1280 (Ref.1) and in ANS-3.1 (Ref.2).

9.2 Methodology

9.2.1 Basic Propositions and Assumptions

In conducting this phase of the project, several propositions and assumptions were tentatively made:

- Consequences resulting from the demonstration of emotionally unstable on-the-job behaviors, in terms of severity and kind, would vary directly as a function of degree of access to vital or protected areas.

- Employees, working regularly in vital or protected areas, and experiencing stress due to the nature of their jobs, would have a relatively high likelihood of exhibiting emotionally unstable on-the-job behaviors.

9.2.2 Approach

The following list includes the basic procedural steps which were followed in evaluating the applicability of our recommended selection procedures for nuclear security personnel to all other nuclear facility personnel:

(1) We reviewed the job descriptions obtained from the representative nuclear facilities referred to in Section 9.1.2, and attempted to identify the critical job characteristics which would justify screening for emotional instability.

(2) We categorized the jobs into the major classifications for nuclear facility positions as outlined in NUREG/CR-1280 and ANS 3.1.

(3) We determined the extent to which indices of emotional instability (see Chapter 3) cut across categories of positions and specific positions.

(4) We judged whether the specific consequences of emotionally unstable on-the-job behaviors would generalize across categories of positions and specific positions with respect to kind and severity.

(5) We determined whether the standards for psychological assessment techniques previously developed (see Chapter 7) would be equally applicable to psychological assessment techniques considered for positions other than security positions.

(6) Finally, we determined whether the instruments recommended as selection procedures for security positions (see Chapter 8) would be equally applicable to all categories of positions, and when not applicable, identified and explained the bases for exceptions to general applicability.

9.3 Results

Based on the methodology discussed in the preceding section, we arrived at the following results:

(1) Indices of emotional instability do generally cut across categories of positions and specific positions. However, the two general categories of emotional instability indices, i.e., 1) inability to appropriately respond to stressful situations, and 2) behaviors indicative of generalized emotional instability, may have different probabilities of appearing among personnel in various positions due to different kinds of job demands.

(2) In some cases, causative factors associated with on-the-job expressions of emotional instability may vary in kind, frequency, and intensity with respect to specific positions.

(3) Specific consequences of emotionally unstable on-the-job behaviors do not always generalize across categories of positions and specific positions with respect to kind and severity. Of considerable importance here is the degree of access and frequency of access with respect to a vital or protected area associated with a given position.

(4) The standards for psychological assessment techniques are applicable to all positions.

(5) The instruments recommended as selection procedures for security positions would generally be appropriate for other positions, with some noted exceptions which immediately follow.*

(6) Some positions, because of the demands associated with them, and given a lack of direct contact with or access to vital areas, would not require any form of screening for emotional instability. These positions would include:

* Transient workers should be subject to the same selection procedures as applicants for permanent employment at a nuclear facility. Since the primary objective of the procedure is the safe operation of the facility, the specific position in question should be the major concern, rather than the prospective employment status of the applicant.

(a) clerical personnel

(b) administrative personnel

(c) accounting and budgetary personnel

(d) public affairs personnel

(e) corporate staff services personnel

(f) real estate and office services personnel

(g) quality assurance personnel who are involved strictly in record-keeping rather than actual inspections

(h) maintenance personnel who would not need to be licensed reactor technicians such as

- cleaners
- painters
- insulation craft workers
- carpenters
- on-site craft workers whose duties are performed only in locations such as the machine shop

(7) Some positions, because of the demands associated with them, would require careful screening pertaining to emotional instability. However, for some of these positions, the specific nature of the position and the conditions surrounding the specific position would not significantly contribute to the expression of emotional instability on the job. That is, for individuals within these positions, factors beyond the specific elements of the job would most likely be the contributing factors associated with the expression of emotional instability. Therefore, for these positions, there would be no need to include in the selection system, simulations which typically would be used to gauge those specific elements of the position critical to the expression of emotional instability. The recommended selection system for these positions would include the MMPI, 16PF, and clinical interview. Positions falling into this selection system would include:

(a) Maintenance personnel authorized to perform work on reactor systems which may affect reactor safety or lead to possible release of radioactivity to the environment. These positions include:

- electronics technicians
- radiation technicians
- chemistry technicians
- test technicians
- quality control inspectors
- machinists
- electricians
- welders
- pipefitters
- grinders
- lead burners
- riggers
- sheet metal workers
- burners/chippers
- maintenance workers such as carpenters, painters, cleaners, and any assistants who would be authorized to perform work on reactor systems.

(b) shift supervisors

(c) plant managers who are strictly responsible for managing and directing their subordinates, rather than holding responsibility for the safe operation and maintenance of the facility as a whole.

(8) Some positions, due to the demands associated with them, would require careful screening pertaining to emotional instability. These positions, because of their specific nature (e.g., stressful demands), and the conditions surrounding them (e.g., constant location in vital or protected areas), would require a selection system which would incorporate the use of simulations. These simulations would need to be developed specifically for each particular position to approximate the specific elements of, and the conditions surrounding the positions, which would contribute to the manifestation of emotionally unstable on-the-job behaviors. Other components of this selection system would consist of the MMPI, the 16PF, and the clinical interview. Positions falling into this selection system would include:

(a) Operators - these individuals manipulate control of a facility or are involved in directing others. This category applies only to those employees who are stationed in the control room and who actually operate or direct the operation of the reactor plant console. Thus, these individuals have the most direct bearing on the status of the reactor. This category includes the following positions:

- senior reactor operators
- licensed reactor operators
- personnel in training for licensing as reactor operators who may be directed to manipulate the equipment
- auxiliary operators
- station attendants

(b) Senior On-Site Managers - this category includes all employees who are senior utility managers at reactor sites and who are in charge of the safe operation and maintenance of the facility.

9.4 Guidelines for Local Validation Studies on Specific Instruments

The following procedural steps are recommended as general guidelines on how a user can validate emotional instability assessment techniques to specific occupations/positions:

(1) On-the-job criterion measures, to be incorporated into subsequent performance appraisal procedures, should be developed for each specific occupation/position. Supervisors should be trained to observe, measure, and document these criteria with respect to their subordinates.

(2) Assessment techniques designed to measure the criterion measures should be administered to employees and/or applicants.

(3) Periodic performance appraisals designed to evaluate the criterion measures discussed in step (1) should be gathered on the aforementioned employees/applicants.

(4) Correlational analyses comparing the results of the predictors with the results of the performance appraisals should be carried out to determine the criterion-oriented validity of the given selection procedure.

(5) See Chapter 10 for a more detailed statement of research strategies.

9.5 References - Chapter 9

1. W. Wegner, Basic Energy Technology Associates, Inc., "Power Plant Staffing," USNRC Report NUREG/CR-1280, March 1980. Available for purchase from National Technical Information Service, Springfield, Virginia 22161.*

2. "American National Standard for Selection and Training of Nuclear Power Plant Personnel," ANSI/ANS-3.1-1978. Available from American National Standards Institute, 1430 Broadway, New York, NY 10018, Copyrighted.

*Also available for purchase from the NRC/GPO Sales Program, U.S. Nuclear Regulatory Commission, Washington, DC 20555.

10. RECOMMENDED RESEARCH

10.1 Introduction

The following recommendations for future research are suggested. It must be stressed that the need for such research is critical given the dearth of available research pertaining to the appropriateness of various instruments for the measurement of emotional instability, particularly in the context of the selection of personnel within the nuclear industry.

10.2 Specific Recommendations

(1) Criterion-oriented validity studies should be carried out to provide evidence for the relationships between various predictors of emotional instability and behavioral indices of emotional instability on the job. In this regard, the following points are stressed:

- As it is recognized that emotional instability is multidimensional, multiple regression research should be initiated to determine the unique contribution of a given predictor to the prediction of emotional instability.

- Effects of moderator variables, such as race and sex, should be investigated to determine the effects of such variables on criterion-oriented validity coefficients. Studies should also be carried out separately for each focal position to determine potential differences in validity coefficients.

- It is recognized that before initiating criterion-oriented validity studies, it may be necessary to better define and develop criteria measures pertaining to the on-the-job indices of emotional instability. Of relevance here is the concurrent contract sponsored by NRC on behavioral observation methods of detecting on-the-job indices of emotional instability* (see section 3.2 for a discussion and listing of these indices).

*This contract has been conducted by Personnel Decisions, Inc., Minneapolis, Minnesota, under NRC Contract Number NRC-01-79-003.

- To place criterion-oriented validity studies in the proper perspective, it is critical that the NRC ensure that the collection of such criteria data be used solely for the research proposed here, as opposed to other uses including auditing purposes.

(2) Content validation strategies should be developed and initiated for those instruments whose basis is purported to be one of content validity. Such strategies should include task congruency analyses indicating overlap between task elements of the job and task elements associated with the selection instrument(s).

(3) Construct validation strategies should be developed and initiated for those instruments whose basis is purported to be one of construct validity. Such studies should include examination of convergent and discriminant validity.

(4) Studies should be initiated to determine factors associated with false positives, i.e. those persons deemed to be suitable for hire, who after job placement, are determined to be emotionally unstable, based on on-the-job criteria. Implementation of such studies would presuppose the existence of measurable on-the-job criteria, or would require the development of measures of such criteria. This latter point has been previously discussed.

(5) Data should be collected pertaining to the inter-rater reliability of the professional judgments made, which are based on a combination of the measures administered to the applicant. Such investigations should focus on the degree to which two or more professionals, combining the same data, arrive at the same conclusions with respect to a sample of applicants. Such studies should focus on:

- All decisions made, that is, accept and reject decisions.
- Reject decisions as a separate subject of investigation.

(6) Data should be collected pertaining to the intra-rater reliability of the professional judgments made, which are based on a combination of the measures administered to the applicant. Such investigations would focus on the degree to which one professional, combining data on a group of applicants, on more than one occasion, arrives at the same conclusions on the various occasions.

The following research, while only indirectly related to this contract, would prove to be extremely informative.

(7) Studies of the effects of various career path opportunities on the composition of the applicant pool should be implemented.

(8) Studies of the effects of various compensation plans on the composition of the applicant pool should be initiated.

(9) Studies of the effects of various compensation plans on the incidence of emotionally unstable on-the-job behaviors should be carried out.

(10) Studies of the effects of job rotation plans on the incidence of emotionally unstable on-the-job behaviors should be initiated.

11. CONSIDERATIONS/CONCLUSIONS/RECOMMENDATIONS

The following subsections are concerned with a variety of issues, some of which have implications beyond simply the selection of nuclear facility personnel. That is, the issues discussed not only focus on special considerations with respect to the selection question, but also on other factors relating to successfully integrating the various functions within the overall organizational structure. The subsections will discuss issues pertaining to personnel/organizational practices, application of standards, evaluators, applicants, and additional assessment considerations.

11.1 Issues Associated with Personnel/Organization Practices

11.1.1 Recruiting Considerations

Regardless of the precision of a selection process, without a desirable applicant pool, selection problems will surface. Thus, a given organization must have a commitment to implementing recruiting procedures, when necessary, to ensure an acceptable work-force. Given the demographics frequently associated with the location of a nuclear facility, i.e., relatively sparse population areas, sufficient numbers of qualified applicants may not exist. This may necessitate the following recruiting steps:

- Recruiting, in general, outside the local area
- Visits to schools outside the local area
- Obtaining of advertising space in media sources outside the local area
- Advertising within the organization's newsletter to obtain "internal" applicants who otherwise would not have demonstrated interest in the security job

11.1.2 Shaping the Work Environment

While one undeniable aspect of some of the jobs in a nuclear facility is the monotonous pace (see Appendix A), there may be other options to the organization other than focusing exclusively on endeavors aimed at measuring an applicant's ability to cope with boredom. Given that the ability to cope with boredom prior to selection may not necessarily be totally predictive of coping with respect to cumulative effects of boredom after selection, organizations may wish to consider means of re-shaping the environment. This might take the form of job rotation where the person experiences some variation in job responsibilities; rotating shifts where novelty is encouraged through the variation in work schedules; train-

ing which would have the dual purpose of enhancing skill level and providing a deviation from the normal work routine; and special provisions for maintaining alertness during periods of reduced psychophysiological arousal. An additional means of confronting the boredom issue would require a realistic depiction of the job at the time of hire. The new employee should be given a realistic orientation to the job through various means, including video-tape presentations, so as not to create false expectations regarding the job.

11.1.3 Enhancing the Attractiveness of Nuclear Facility Positions

Given that dissatisfaction with the job may often lead to turnover (Refs. 1, 2, and 3), and/or unwillingness to tolerate job-related stress, some consideration should be given to enhancing the attractiveness of some nuclear facility positions. Several suggestions immediately follow:

- Efforts should be made to improve wages and benefits associated with given positions. Including certain positions within the management compensation program, for example, would also indirectly elevate the status of these positions.

- Attempts should be made to identify logical career paths for position incumbents. This might encourage integrating different functions within a given career path, creating job families, etc.

11.1.4 Self-Control Procedures to Manage Stress

It is clear from the research, that for many positions, some degree of stress is unavoidable. Thus, in addition to selecting persons who are most capable of coping with stress, consideration should be given to enrolling incumbents in stress management training programs.

11.2 Special Issues Regarding the Application of Standards

11.2.1 Internal vs. External Security Forces

The standards of selection for security personnel should be uniformly applied whether the hiring organization relies upon an internal or external security force. If an external force is to be used, it is incumbent upon the hiring organization to ensure that the vendor is adhering to proper selection standards. Dependence upon an external force should not in any way detract from the hiring organization's allegiance to proper selection standards.

11.2.2 Job Applicants from within the Organization

Quite often, at a particular point in time, persons who are currently employed by the organization in a non-nuclear site position may wish to be considered for a nuclear site position. The question then becomes: How shall this person's application be treated, since he/she has already been an employee within the organization for some period of time?

Assuming that standards and procedures would exist with respect to screening out individuals who are emotionally unstable and would be risks as nuclear facility personnel, it would appear logical that the same standards be applied to internal applicants. Simply because one has been an accepted employee in a job within the organization, does not necessarily mean that this individual will be able to cope with the quite different pressures and demands associated with the nuclear facility position.

One exception to the same standard and procedures principle would be in order. If the organization has instituted a behavioral observation program (see Section 10.2), used to observe indices of emotional instability on the job, and the person's present job is in the category of jobs falling into the behavioral observation program, then data would be available attesting to the emotional stability of this individual. It should be stressed that this behavioral observation program should have to meet the requirements promulgated in this review regarding such psychometric qualities of reliability and validity, and that supervisors within the organization implementing the behavioral observation program be properly trained in the skill of observation.

In addition, it is emphasized that the substitution of the behavioral observation program for selection procedures otherwise used may only be partial. That is, behavioral observation carried out on an unrelated job may provide only a partial picture as to whether or not the individual in question would perform in a stable manner with regard to the nuclear facility position. For example, merely because a secretary is observed by his/her supervisor over a twelve month period as not having expressed any unstable behaviors on the job, does not predict completely the presence of such stability when placed in a nuclear facility position. Certainly, we would not select nuclear facility personnel on the basis of ability demonstrated as a secretary, and likewise, this would also be the case when focusing exclusively on emotional instability.

11.2.3 Application to the Current Work-Force

Given that industry-wide standards for selection exist, the question needing to be addressed focuses on the issue of retroactive application of the standards to the current workforce. Two options appear to be available with respect to the issue of retroactive application of the standards. First, if an acceptable continual behavioral observation program currently exists, meeting the measurement principles promulgated throughout this review, and provides data regarding a given employee's record of emotional stability as expressed on the job, it would appear that there would be no need to retroactively apply the selection standards.

However, if no such behavioral observation program is in operation, and if the organization's selection procedures are found to be lacking in accordance with the selection standards, it would seem to be in order to apply the selection standards retroactively with respect to given individuals.

11.2.4 Need for Periodic Reevaluation and On-the-Job Observation

The need for periodic reevaluation must be considered within the context of the cumulative effects of stress and/or boredom for the individual in the nuclear facility position. Simply because an individual, at the time of hire, is deemed emotionally stable, does not automatically guarantee continued emotional stability. Thus, consideration must be given to the need of periodic reevaluation to monitor possible decrements in stability.

However, given the implementation or presence of a carefully developed behavioral observation program, meeting the psychometric principles discussed in this review, periodic reevaluation would not be necessary.

11.2.5 Nature of the Hiring Decision

The hiring decision may not be a simple yes or no. Alternative decisions would include: an accept decision, a non-accept decision, and an accept decision qualified by developmental recommendations. In the latter case, guidance should be provided to the prospective supervisor pertaining to areas of concern detected during the evaluation process, which should be monitored by the supervisor through a behavioral observation on-the-job program. These areas of concern, revealed through the evaluation process, should be specified in behavioral terms, as opposed to clinical terms. Examples of such behavioral areas of concern would include questions regarding appropriate interactions with peers, interactions with superiors, etc.

Only information such as that specified above, germane to the position under consideration, should be released to the prospective supervisor.

Before the hiring decision is made, all information under review, including background data, technical competence information, etc., should be integrated and examined in combined form.

Written and oral feedback should be provided to the evaluatee indicating the reasons for rejection. Such feedback should be specific and behaviorally oriented, e.g., difficulty relating to peers. (See Section 11.4.1 for further discussion of the nature of feedback to applicants).

11.2.6 Administration and Control of the Evaluation Process - Where in the Organization Should the Responsibility Reside for Retention of Records?

Which entity in the organization should have the responsibility for administering and controlling the evaluation process? Potential areas of responsibility would include Personnel, Training Unit, Health Services, the Functional Unit where the position is located, and Employee Assistance.

While each of these organizational entities may clearly be involved in a segment of the evaluation process, overall responsibility should be placed entirely within one entity to ensure uniformity and proper control of evaluation procedures.

In determining the choice of organizational entity, the following factors should be considered:

- The presence of systems which will ensure the confidentiality of records
- The history and previous success in maintaining the confidentiality of records. The location of the counseling unit
- Experience in the objective interpretation of data
- Existing state legislation pertaining to the confidentiality of personnel records
- Where the hiring decision is ultimately made
- Where data used for the hiring decision are obtained

- Where there are personnel with experience in translating professional/technical terms into lay-person language

- Where hiring data can be most easily compared to subsequent on-the-job performance data

- Where existing government legislation of a non-selection nature does not compromise the confidentiality of records, and allows for the inspection of such records.

- The entity in the organization where first contact with the job applicant is made

- Experience in supporting the accuracy of data used for hiring in situations where hiring decisions are grieved?

11.2.7 Procedure to be Followed if Faking is Detected

One of the individual instrument standards outlined in Chapter 7 deals with an instrument's susceptibility to faking. This standard assumes that an ideal instrument would be able to detect an applicant's intentional distortion of results. In the context of a job application process, it might be expected that an applicant wanting a job would attempt to make him/herself look as good as possible, even if some distortion of data is involved. Since this types of occurrence might be considered a "normal" tendency, the following procedure is recommended when faking is detected:

(1) The professional who conducts the clinical interview should explain to the applicant whose test results have indicated faking that while it is understood that he/she is trying to make a good impression, factual information is necessary.

(2) If the applicant's distortion is minimal, the discrepancies may be clarified by direct questioning during the interview.

(3) If it appears that the results have been so distorted that they are completely invalid, the applicant should be asked to submit to an additional administration of the instrument(s) in question, with instructions to respond as truthfully as possible during the second administration.

(4) If the results of the second administration still indicate intentional distortion of results, the professional should recommend that the applicant be rejected on the basis of intentional deception.

11.2.8 Implementation of Standards - Procedures to be Followed by the NRC and the User Organizations

The following implementation steps are recommended.

(1) Choice 1: Organization uses system that has been recommended.

(a) Representative from organization fills out Form A (see appendix F) indicating instruments to be used without any further justification.

(2) Choice 2: Organization desires to use instrument(s) different from those recommended

(a) Representative from organization fills out Form B indicating each instrument desired for use and includes a justification statement with respect to each prescribed standard.

(b) NRC expert, a psychologist, meeting professional standards, reviews the form submitted by the organization and determines whether instruments are acceptable in accordance with standards. If not, the NRC professional responds to the organization by asking for more justification with Form B, indicating areas (standards) where more justification is needed.

(c) At this point, the organization can choose to respond to Form B. If the organization does not respond, it is not permitted to use the instrument(s) requested. If the organization does respond to Form B by submitting additional justification, the NRC psychologist reviews it and informs the organization of an accept or reject decision.

(d) If the organization receives a reject decision, it must re-submit a proposal for substitute instrument(s) and go through the review process again.

All NRC responses to organizational requests should be made in standard written form.

11.3 Issues Pertaining to the Evaluators

11.3.1 Qualifications and Standards of the Professional

This is a very critical question, for without having qualified professionals, any testing program is likely to experience failure. Standards for professionals involved in the evaluation process should include the following elements:

- Thorough training in observational techniques used to measure individual differences in human behavior
- Exposure to the industrial/organizational structure of the nuclear facility setting
- Knowledge of and substantial experience with psychometric and other measurement procedures applied to an organizational or industrial environment
- Working knowledge of the cultural/regional environment of the work-force
- Adherence to standards endorsed by the national professional group(s) with which professional is affiliated
- Either a (1) state-licensed doctoral level psychologist with a Ph.D. in clinical, counseling, or industrial psychology from an American Psychological Association approved program, or (2) a board certified or board eligible psychiatrist
- At least five years (including a minimum of two years post-doctoral) experience in an individual assessment context
- Exposure to orientation program offered by organization to all new employees working at a nuclear facility
- In addition to the aforementioned orientation, professionals should experience an additional orientation program relating to security provisions of the facility, regulations relating to record keeping, and a familiarity with the job(s) including available job analysis results pertaining to given positions.

11.3.2 Professionals: Internal vs. External with Respect to the Utility Organization

While advocacy of internal or external professionals is not put forth, the following considerations are presented:

(1) Advantages of Internal Professionals

- Internal professionals may be in a position to better integrate selection data with other on-going organizational systems such as recruitment and training
- Internal professionals may have better access to behavioral on-the-job criteria relating to specific indices of emotional instability
- Costs to the utility organization may be lessened by having on staff, full-time internal professionals
- Availability of professional may be enhanced by having on staff, full-time internal professional
- Feedback to unsuccessful applicant may be more conveniently scheduled given availability of full-time internal professional
- Internal professional may have a better appreciation of the organizational climate

(2) Advantages of External Professional

- Feedback to unsuccessful applicant may be more constructive given applicants' presumed greater willingness to be more open to external professional.
- External professional might appear to be more objective given less full-time reliance to the "client" organization.
- External professional may have broader range of knowledge given more experience with a variety of nuclear industry organizations.
- External professional may be able to provide more confidentiality as he/she resides outside the organization.
- External professional might be more knowledgable of remedial services (e.g., counseling) locally available to applicant.

11.4 Issues Pertaining to the Applicants

11.4.1 Nature of the Feedback

Of critical importance is the feedback which is received by the applicant who has not passed the evaluation process. The following considerations need to be taken into account:

- Feedback should be of an oral and written nature. Given the seriousness of the selection decision and the basis of rejection, i.e., emotional instability, a thorough statement of the basis for the selection decision should be provided to the unsuccessful applicant.

- While thorough de-briefing might heighten the applicant's awareness of his/her state of emotional instability, this can be somewhat reduced by placing the discussion within the context of the given position's work demands and environment. That is, the unsuccessful applicant should not have the misconception that he/she is emotionally unstable, but that his/her chances of coping with the degree of stress present in the position are somewhat limited. Specificity of information is a necessity to guard against misunderstandings.

- Oral feedback should be provided by a highly trained professional, i.e. psychologist or psychiatrist.

- Request for information regarding means of improving oneself for purposes of re-examination should be dealt with through a carefully worded written policy statement. This would include a list of local professionals in the counseling field, if the applicant indicated a desire for assistance in improving his/her level of emotional stability.

11.4.2 Rejection and Exclusion From Other Fields

Rejection from a given position based on evidence pertaining to emotional instability should not automatically be used as a basis for exclusion from other positions within the organization. Of critical importance in determining suitability for other positions would be the degree of correspondence between the demands of the given position and the other positions desired by the individual.

11.4.3 Policy of Re-examination

In accordance with Equal Employment Opportunity Guidelines of 1970, provisions should be made for the re-testing of applicants after some period of time after initial rejection. Policies need to be established, and should take into account:

- The measurement approach used to gauge emotional instability.
- The theoretical assumptions pertaining to the relative permanency of a state of emotional instability.
- The life of the data obtained through the selection process.

11.4.4 Appellate Process

A formal appellate process should be established for the purpose of resolving an applicant's disagreement with the hiring decision. Any appellate system must be characterized by objectivity. The following elements should be addressed in the appellate process:

- The person being evaluated should be entitled to disclosure of the testing results pertaining to his/her evaluation.
- The appellate process should be in the form of a written policy, to which no deviation is allowed, thus ensuring uniformity in treatment of all evaluatees.
- The appellate process should define the contact person within the organization with whom the process is initiated.
- Time parameters associated with the appellate process must be defined, and reasonable, so as to prevent long, drawn-out appeals, which would prove detrimental and costly to the evaluatee.
- To appeal the professional evaluation, the evaluatee may choose a second professional of his/her own choice or may follow recommendations of the hiring organization, but in all cases the professional must meet the professional standards addressed in this document.
- Evaluations subsequent to the initial evaluation should utilize the same criteria for the hiring decision as employed in the first evaluation.

- Evaluations subsequent to the initial evaluation must be independent of the initial evaluations, and it is the hiring organization's responsibility to ensure that such "blind" evaluations are carried out.

- To further guarantee objectivity, evaluations subsequent to the initial evaluation, should be conducted by professionals who are not full-time employees of the hiring organization.

- If the subsequent evaluation results yield a "tie" decision, a third professional should be called upon to review the data developed by the first two evaluations, and make the final recommendation.

- All data involved in the appellate process must be treated as a confidential matter.

11.5 Issues Pertaining to Additional Assessment Considerations

11.5.1 Issue of Literacy

If written materials are to be used during the selection procedure, they should not be so difficult as to discriminate against certain sub-groups in the population. Literacy tests should be considered as a source of information to determine the extent to which unrelated, non-job written requirements preclude entry into the job. In such cases, other means of presenting the material, e.g., video-tape, should be considered.

11.5.2 Fitness Beyond the Job

Among the issues of serious concern is that centering on the assumption of fitness beyond the job. While adherence to this assumption is not explicit, and is certainly not restricted to the selection procedures within the nuclear industry field, its effects in conjunction with the environment surrounding a given nuclear facility position may be profound and unique.

More specifically, it appears that selection requirements for nuclear facility positions on occasion are too severe given the requirements of the job. Requiring, for example, applicants for certain given positions to have college educations, as some utility concerns do, may be excessive with respect to the demands of the job, and in fact, is somewhat at odds with legal rulings (Griggs vs. Duke Power Co., 1971).

The effects of this selection orientation may be significant when viewed within the context of the actual requirements of the job, and the environment in which the job is placed. If,

in fact, the person with the college education finds him/herself in a situation where his/her skills/education are under-utilized, considerable dissatisfaction may set in, (Locke, 1976) and coupled with the usual monotonous aspects of the given position, may, over a period of time, increase the likelihood of manifestation of emotionally unstable behaviors expressed on the job.

It is important, therefore, for any nuclear facility, in the development or revamping of its selection procedures for given positions, to seriously consider the extent to which the fitness beyond job assumption is being applied, even in the most subtle form, with respect to any of its selection procedures.

11.5.3 NRC and EEOC Considerations

As Uhrig (1979) pointed out, "recent court decisions indicate that any requirements for hiring a person must be directly job-related. ... Clear guidance must be provided to the industry resolving any potential conflicts between the clearance program and the equal employment opportunity laws."

It is essential then that some guidance be provided to the utility organizations pertaining to how NRC requirements are to interface with EEOC standards.

EEOC statements distinguish one's personal life from his/her job, and would preclude quite often, for example, investigations into the person's past off-the-job behaviors. Yet, the NRC would argue quite strenuously that it is appropriate to investigate one's credit history, for example, to determine bad risk (unstable) elements if one is being evaluated for example, for a nuclear facility position.

Related to the above, it appears that EEOC focuses its orientation to selection on normal requirements and events associated with the job, e.g., checking the badges of employees, while NRC through its selection perspective, seems to be concerned about protection with respect to unusual events that might occur, e.g., infiltration of a facility by outside agitators. Of critical importance in this regard is a consensus on the definition of critical or important work behaviors. In discussing the basis for determination of critical or important work behaviors, the Uniform Guidelines (1978) state:

> The report should describe the basis on which the behavior(s) or outcome(s) were determined to be critical or important, such as the proportion of time spent on therespective behaviors, their level of difficulty, their frequency of performance, the consequences of error, or other appropriate factors. (15B[3]

The consequences of error appears to be the only one of these considerations which justifies appropriate behavior in crisis situations as being an important work behavior.

It is not the purpose of this section to resolve or define the interface between NRC and EEOC, as this ultimately will be left up to the courts, but a few questions are offered which need to be studied.

First, the question must be answered - to what extent does the potential for a catastrophic event allow for the denial of some individual rights within the context of selection? Second, to what degree are we able to establish a means for identifying the probability that such a catastrophic event can or will occur?

11.5.4 Indices of Emotional Instability and Organic Brain Syndromes

Several of the behaviors that we have used as indices of emotional instability, such as slurred speech, poor psychomotor coordination, memory loss, and delayed reaction time, may be indicative of a group of physiological conditions collectively known as organic brain syndrome. Therefore, if such behaviors are observed during the selection procedure, the applicant involved should be recommended for further medical screening prior to consideration for rejection on the basis of emotional instability.

11.5.5 Selection Issues and Chemical Substances

Another issue of major concern is the effect of drugs and alcohol on employees' on-the-job performance. In this respect, concern must be given to the effects of both prescribed and nonprescribed drugs.

One broad group of drugs may affect mental performance and decision-making on the job -- the psychotropics or mind-affecting drugs. These include such sub-groups as Central Nervous System (CNS) stimulants, antidepressants, sedatives, anesthetics, and tranquilizers -- any of which may be prescribed for legitimate medical reasons. The same drugs plus illegal substances such as marijuana and LSD are subject to misuse.

Other drugs not classified or used as psychotropics may have at least some CNS action as a side effect to a completely different desired effect. Some anti-hypertensives, for example, could be listed here. For purposes of this discussion the term psychotropic will cover all drugs which affect the mind as either a desired or side effect.

Even though a psychotropic drug has been prescribed by the personal physician and is being taken correctly, its possible effect on the mental clarity and reaction time of the individual faced with a sudden emergency must be seriously considered. Therefore the following recommendations are made:

- A complete and accurate medical and drug history must be obtained during the evaluation process.
- Individuals on psychotropic drugs at time of evaluation should not be hired.
- After hiring, if a psychotropic drug is prescribed by the individual's personal physician, it must be reported immediately to the organization or organization physician. The employee should be transferred to non-critical work until the drug treatment is finished.
- Failure to completely disclose medical and drug history at time of evaluation, or in the use of prescribed drugs during employment, or any use of hallucinogens, should be considered as grounds for dismissal or suspension.
- Alcohol is a special case of a psychotropic drug. Working while under the influence of alcohol, or alcoholism as a chronic condition, requires removal to non-critical areas at the very least.

11.5.6 Additional Cautions

Attempts should be made to avoid selecting persons who have a distorted impression of the job, i.e., that opportunity exists for unlimited adventure. Information provided during the evaluation process should enable the applicant to see that unlimited adventure is not associated with the job, and perhaps he/she needs to reconsider his/her aspirations.

Previous law enforcement or military experience should not be assumed as a prerequisite for hire. In fact, such experience might foster the development of a false expectation on the part of the applicant, which might lead to frustration subsequent to selection.

11.6 References - Section 10.

1. E.E. Lawler, III, "Job Attitudes and Employee Motivation: Theory, Research, and Practice," Personnel Psychology, 23, 233-237 (1970). Available in public technical libraries.

2. L.W Porter and R.M. Steers, "Organizational, Work, and Personal Factors in Employee Turnover and Absenteeism," Psychological Bulletin, 80, 151-176 (1973). Available in public technical libraries.

3. V. Vroom, Work and Motivation, Wiley, New York, 1964.

APPENDIX A

NUCLEAR REGULATORY COMMISSION

Job Analysis Report

Position:
Security Personnel

Submitted by:
Assessment Designs, Inc.

Under Contract Number NRC-01-79-002

CONTENTS

1. INTRODUCTION

The pages which follow comprise the job analysis of Nuclear Facility Security Personnel. The purpose of the job analysis was to identify the critical demands and activities of the security jobs. A thorough understanding of a job is essential prior to determining the appropriateness of potential selection systems. Both supervisory and non-supervisory positions were included in the job analysis.

Also considered is the issue of emotional stability and its relationship to the performance of security personnel. Factors of the job which may affect the emotional stability/performance of security personnel and behavioral patterns which may be indicative of inadequate performance and emotional instability are also discussed in the job analysis.

2. METHODOLOGY

2.1 Sources of Information

- Basic information on all security positions was reviewed.
- Existing job analyses information compiled by independent nuclear related organizations both in government and industry were reviewed.
- Discussions were held with personnel in the nuclear industry, both in the United States and abroad. These persons came from the following job categories:

  Security guards

  Security supervisors

  Non-security management personnel

  Personnel administrators

  Training personnel

  In-house psychologists

2.2 Interview Format

When discussions were carried out, the following types of questions were asked:

- Is the security force internal or external?
- What are the different types of security positions?
- Do current job analyses exist?
- What is the average tenure of security personnel?
- What are the work hours? Are the work hours flexible?
- How many work shifts are there? Is shift rotation used?
- How many employees are in the security department? What positions do they work in?
- What are the major functions, responsibilities, and activities required of security personnel?
- What skills/abilities are required for effective performance?

- What areas/activities differentiate effective performance from ineffective performance?
- What areas/activities do incumbents currently have the most difficulty in?
- Which, if any, characteristics, activities, and responsibilities of the job are perceived as undesirable by the incumbents?
- Is job/task rotation used?
- What problems are perceived as 'real' by the security personnel (e.g. intruders, equipment malfunction)?
- What are the promotional/career opportunities?
- What types (and examples) of undesirable behavior may occur on the job?
- What is currently being done with respect to the selection of individuals for security positions? How does this differ from selection procedures used for other employees within the organization?

3. SECURITY PERSONNEL - DEFINITION

3.1 Position Delineation

At various nuclear facilities, and often within a single nuclear facility, the formal position title of security guard is used to refer to different, but similar, jobs. Two security guards may not perform exactly the same tasks and may not be responsible for the same duties. Occasionally, two security employees who have different formal job titles may actually have the same responsibilities and perform the same tasks. For clarification purposes, the following job titles will be used when referring to security employees:

Job Title	Job Description
Security Guard	A security employee who does not supervise other security employees
Security Supervisor	A security employee who does supervise other security employees
Security Personnel	All security employees, both guards and supervisors

3.2 Examples of Frequently Used Job Titles

- Security Guard
- Security Officer
- Access Control Officer
- Public Safety Officer

4. JOB TASKS - JOB ACTIVITIES - KEY FACTORS

4.1 Job Tasks - Job Activities

One purpose of the job analysis was to determine the types of actual tasks required of security personnel. While the absolute number of tasks required is quite large, the analysis was concerned with identifying the major job tasks performed. For this reason, the voluminous information acquired was reduced to reflect the major tasks associated with the positions. Table 1 presents the major tasks associated with both the security guard and security supervisor positions, while Table 2 presents additional major tasks associated with only the security supervisor position. Table 1 includes examples of specific job activities associated with the job task. These job activities appear in the parentheses following the job task.

4.2 Key Factors of the Job

Consideration was given to the relationship between the job tasks and indices of emotional stability. Key factors are those aspects of the job task which may, in and of themselves, affect the performance/emotional stability of the security personnel performing the task. The following key factors have been used within the context of discussing indices of emotional stability:

Key Factors	Definition
Boredom	An uninteresting task; a non-challenging task.
Monotony	Performing the same, or similar, activities continuously; low diversification of activities.
Possibility of Bodily Injury	Perceiving that performing a task may cause personal injury to oneself and/or others.

Possibility of Property Damage	Perceiving that performing a task may lead to damage of property (e.g. nuclear facility, personal possessions located in the facility, automobiles or homes in the vicinity).
Adverse Work/Weather Conditions -	Performing a task in undesirable conditions (e.g. cold weather; snow; sleet; stationed in a small area).

Performing a boring or monotonous task can cause a decrease in the mental and physical alertness of security personnel, which, in turn, may cause a decrease in the employee's level of performance. For this reason, boredom and monotony are considered to be key factors of a job task.

Security personnel who perceive that performing a particular task may cause either personal injury to themselves or others, or cause damage to personal, private, or public property, may be reluctant to perform that task. It is important in this regard to realize that what matters is how the individual performing the task perceives the task and not how another employee or anyone else, for that matter, perceives it.

Security personnel may also be reluctant to perform job tasks which require them to be exposed to adverse work or weather conditions.

Reluctance to perform a task can lead to procrastination in performing the task, (e.g. a security guard hesitates before investigating the presence of an intruder), incomplete or inadequate performance of a task (e.g. security guard assigned outside foot patrol on an extremely cold day may only complete part of his/her assigned patrol), or non-performance of the task (e.g. a security guard decides not to investigate the presence an unidentified package). Thus, the possibility of bodily harm, the possibility of property damage, and adverse work/weather conditions are considered to be key factors of a job task.

Where appropriate, in Table 1, these key factors are included with respect to the description of the job task.

TABLE 1

Major Job Tasks - Job Activities - Key Factors

Position of Security Personnel

- Monitor central security alarm system (closed circuit television system, access control system, intrusion detection system, fire alarm system), identify and react to abnormal readings and/or alarm activation (notify intruder detainment squad, reduce or eliminate access to specific areas, personally relocate to a specific area, notify local law enforcement agency, notify fire control personnel).

Key Factors

Boredom - During periods of normal to low activity levels this task may be uninteresting. Since the task requires constant monitoring of the system, security personnel are not allowed to move about readily when performing this task.

Monotony - Low diversification of daily job activities.

- Monitor the central security alarm back-up system (identical to the central security alarm system except that it is only used during emergencies and when the main system is inoperable).

Key Factors

Boredom - Since this is only a back-up system the security person may assume that another security person will be constantly monitoring the same sensory inputs. For this reason, performing this task has less importance associated with it than monitoring the main system and also has more boredom associated with it.

Monotony - Low diversification of daily activities.

- Control entry and exit of nuclear facility personnel (verify employee identification, assist in the distribution of employee identification badges, conduct search of physical packages, conduct metal detector search).

Key Factors

Boredom - During time periods other than shift changes this is a low activity level task. Few employees enter or exit the nuclear facility except during shift change.

Monotony -	Performing the same activities continuously.
Possibility of Bodily Injury -	An irate employee, or an intruder posing as an employee may inflict bodily harm upon the security employee or others.
Possibility of Property Damage	An intruder masquerading as an employee may possess a package that contains explosives.
Adverse Work/ Weather Conditions -	May be stationed in a small, isolated area. May be stationed outdoors, susceptible to undesirable weather conditions.

- Control entry and exit of nuclear facility visitors (verify identification of visitor, ascertain purpose of visit, issue appropriate visitor security badge, conduct metal detector search, conduct search of physical packages, notify appropriate employees of visitor's arrival, escort visitor to desired location, provide information to visitor, maintain identification badges).

Key Factors

Boredom -	During times of few visitors this task will entail few job activities.
Possibility of Bodily Injury -	An unfriendly visitor may inflict bodily harm upon the security employee or other persons.
Possibility of Property Damage -	An unfriendly visitor may possess a package containing explosives.
Adverse work/ Weather Conditions -	May be stationed in a small, isolated area. May be stationed outdoors, susceptible to undesirable weather conditions.

- Control entry and exit of vehicles (verify identification of vehicle and driver, conduct search of vehicle and physical packages, escort non-designated vehicles).

Key Factors

Boredom -	May be long and frequent periods of inactivity.
Possibility of Bodily Injury -	Driver and/or passenger in vehicle may be unfriendly. Security employee is at a disadvantage in that he/she may not be able to act quickly and to easily detect the possession of a weapon by a person, or persons, in the vehicle.

Possibility of Property Damage -	Unfriendly persons in the vehicle may possess weapons that will cause damage to property.
Adverse Work/ Weather Conditions -	May be stationed in a small, isolated area. May be required to perform task outdoors, susceptible to undesirable weather conditions.

- Conduct foot and/or vehicle patrol of nuclear facility and grounds for the purposes of: identifying possible security hazards; identifying possible fire hazards; detecting the presence of unauthorized vehicles, packages, persons; detecting unusual activity; detecting equipment in need of repair (e.g. alarms, barricades, fire extinguishers); complete written report detailing patrol.

Key Factors

Monotony -	Perform the same activities continuously. May perform these activities many times before ever being called upon to actually detect something of an unusual nature.
Possibility of Bodily Injury -	Bodily injury may be incurred as a result of faulty equipment, unauthorized packages, or the actions of an intruder.
Possibility of Property Damage -	Property damage may be incurred as a result of faulty equipment, unauthorized packages, or the actions of an intruder.
Adverse Work/Weather Conditions -	Foot patrol may be extensive and strenuous. Outdoor patrol may be required during undesirable weather conditions.

- Conduct search for unusual objects (e.g., bomb); identify object; report existence of object; investigate object; conduct or assist in investigation of how object was placed in the nuclear facility.

Key Factors

Possibility of Bodily Injury -	This element is strongly associated with this task. Security personnel are attuned to the fact that an unidentified object may cause bodily injury to themselves and/or others.
Possibility of Property Damage -	This element is also very much related to this task. An unidentified object may cause property damage.

Adverse Work/Weather Conditions - Task may be performed outdoors in undesirable weather.

- Properly react to the identification of something of an unusual nature (notify central security alarm system operator, detain intruder, correct fire or safety hazard, evacuate the area, isolate the area).

Key Factors

Possibility of Bodily Injury - This element is very much associated with this task. Knowledge of the fact that improper action on the part of the security employee may cause bodily injury to a person or persons places extreme pressure on the security employee.

Possibility of Property Damage - This factor is strongly related to this task. Security personnel realize that improper action on their part may result in extensive property damage.

Adverse Work/Weather Conditions - This task may have to be performed outside in undesirable weather.

- Initiate security alert and/or initiate security alert procedures (reduce access to specific areas, evacuate and isolate specific areas, detain intruders).

Key Factors

Possibility of Bodily Injury - Performance of this task may cause bodily harm to security personnel and others. Improper or delayed action on the part of a security employee may result in personal injury.

Possibility of Property Damage - During performance of this task, it is possible that property will be damaged.

Adverse Work/Weather Conditions - This task may have to be performed outside in undesirable weather.

- Perform emergency medical treatment (attend emergency medical treatment training sessions; transport sick or injured persons; identify injuries; determine what, if any, type of medical care to administer; administer appropriate medical care).

Key Factors

Possibility of Bodily Injury - Administering improper medical care, or improperly administering the correct care, can cause additional harm to the injured person. Also, additional pressure results from the perception that security personnel may be subject to legal action as a result of administering medical treatment.

Adverse Work/Weather Conditions - Task may be performed in undesirable weather and/or emergency (e.g., sabotage, nuclear equipment breakdown) situations.

- Prepare, review, and evaluate written materials (complete required daily, weekly, monthly... reports).

Key Factors

Boredom - This is a relatively unchallenging and uninteresting task that requires thoroughness and attention to detail.

Monotony - This task is performed on a regular basis. Performance of this task, which accounts for a significant percentage of a security employee's work day, is very routine.

- Interact effectively with the public*. (Conduct public education programs, provide information to visitors, represent the nuclear facility to the public.)

Key Factors

Monotony - Security personnel may frequently have to answer the same questions asked by different visitors. However, the diversification of activities, which results from interacting with the public and from performing a task which is dissimilar to those normally performed by security personnel, may reduce the monotony associated with this task.

- Report problems, concerns, and activities honestly and straightforwardly to upper management (determine what information should be communicated upward, inform superiors of irregular behavior of fellow security personnel/peers).

* The security employee may be the only nuclear facility employee with whom the public has contact. Public opinion of the nuclear facility may be influenced by the security employee's performance and demeanor.

Possibility of Bodily Injury - Security personnel may perceive that performance of this task (e.g., reporting incidences of irregular behavior on the part of fellow employees) may result in retaliation by said employee. This retaliation may be in the form of personal injury.

- Make rational, effective, and immediate decisions based on company policy, previous experience, and individual knowledge.

Key Factors

Possibility of Bodily Injury - The results of a bad decision may cause bodily injury to a person or persons.

Possibility of Property Damage - The results of a bad decision may cause damage to property.

Adverse Work/Weather Conditions - This task will frequently be performed during emergency situations. Security personnel must make an almost instantaneous transition from performing in a routine, standardized situation to performing in an emergency, possibly life threatening, situation.

- Handle firearms properly. (Competent in the use of firearms, remove firearm from holster only when necessary.)

Key Factors

Possibility of Bodily Injury - Misuse of a firearm may result in injury to a person or persons. Additionally, being armed may result in an intruder over-reacting, thus causing possible physical harm to the security personnel.

- Interact with people of various physical and mental abilities/disabilities.

Key Factors

Possibility of Bodily Injury - Security personnel may interact with persons who possess a tendency to act in a violent manner.

Possibility of Property Damage - Security personnel may interact with persons who have a tendency to act in a violent manner which may result in damage to nuclear facility property.

TABLE 2

Additional Major Job Tasks

Position of Security Supervisor

- Maintain inventory of supplies and order supplies when needed.
- Design and conduct security guard training programs (make training recommendations, design training programs, develop training criteria, determine training needs, determine training validation and scoring).
- Provide performance feedback to subordinates (annual performance appraisal, informal performance review, counsel subordinates with performance problems).
- Provide guidance to subordinates (assign work activities to subordinates, transmit and interpret company and upper management policies and guidelines to subordinates, delegate work in accordance with the needs of subordinates).
- Prepare work schedule for self and subordinates.
- Prepare, review and evaluate operational reports of total work group as well as those of subordinates (subordinate's time application reports, vehicle usage reports, enforcement activities and results achieved by the total work group).
- Maintain group cohesion and motivation among others (explain reasons behind decisions and actions, interact in a fair and consistent fashion with all subordinates, actively reward good performance by others, show interest in the performance and development of subordinates).
- Prepare budget for the security department and operate within the budget.
- Devise, direct, and conduct security vulnerability tests (conduct simulated intruder attacks, conduct sabotage simulation, evaluate performance of security personnel during simulations, implement procedures to improve weaknesses detected during simulation).

5. SECURITY SUPERVISOR - SECURITY GUARD DISTINCTIONS

5.1 Scope of Responsibility

Security guards are responsible for performance of the activities and tasks required of their own position. At certain times, however, security guards may be required to perform activities that are not routinely necessitated by their specific position (e.g., emergency activities). In either case, security guards are responsible for all activities they do perform, whether these duties are routinely assigned to their position, or are required by unusual circumstances.

Security supervisors, like security guards, are responsible for performance of the activities and tasks required of their own position and for the performance of activities and tasks required during emergency situations. Unlike security guards, however, security supervisors have the additional tasks of coordinating and monitoring the activities and performance of their subordinates.

5.2 Relationship of Responsibility to Performance/Emotional Stability

The additional responsibilities of security supervisors may reduce the amount of boredom and monotony associated with their jobs, while at the same time increasing the amount of perceived stress associated with them. It can be generally assumed that although the demands of the positions of security supervisors and security guards are somewhat different, these demands will affect the performance/ emotional stability of both personnel levels. Additionally since the demands of the two positions differ, stress present on the two jobs, along with the source of the stress, will also vary.

6. JOB CHARACTERISTICS WHICH MAY AFFECT THE PERFORMANCE/ EMOTIONAL STABILITY OF SECURITY PERSONNEL

Aspects of the security employee position, which are not classified as job tasks or job activities, may impact upon the performance/emotional stability of security employees. Listed below are some of these job aspects.

- Employee benefits may be low.
- May work extended shifts (i.e., longer than an eight hour shift).
- Fear of the effects of exposure to radiation (married employees concerned about the effects on pregnancies).

- Career growth within the security department is often limited.

- May be required to work in a small enclosed area (e.g., bullet proof cage) for an entire shift.

- Daily work may be repetitious, not much diversity in job activities.

- Job often ranges from extreme boredom (i.e., working night shift) to job overload situations (i.e., employees are both arriving to and leaving the facility).

7. BEHAVIORS AND BEHAVIORAL PATTERNS ASSOCIATED WITH RISKS TO NUCLEAR FACILITY SECURITY

7.1 Introduction/Explanation

Table 3 presents a list of behaviors and behavioral patterns that would pose a risk to the security of a Nuclear Facility. Any given behavior may vary in terms of five major parameters. It is vital that the following five parameters be considered in the evaluation of a behavior, since they strongly contribute to the determination of whether a behavior is appropriate or inappropriate:

- frequency - how often the behavior occurs within a specified time period.

- intensity - the degree of strength or severity of the behavior.

- duration - the length of time elapsing between the initiation and the termination of the behavior.

- latency - the length of time elapsing between the occurrence of stressful stimulus situations and the appearance of a detrimental impact on the performance/emotional stability of an employee.

- threshold - the employee's individual level of tolerance to stressful or demanding conditions, which, when surpassed, results in the deterioration of performance/emotional stability.

The behaviors shown in Table 3 may occur singularly or in conjunction with one another. It should be noted, however, that these behaviors and behavioral patterns are not to be viewed as clinical syndromes, but as behavioral patterns which may appear on the job.

TABLE 3

Behaviors and Behavioral Patterns Associated

With Risks to Nuclear Facility Security

A. Behavioral Patterns Indicative of Inability to Appropriately Respond to Stressful/Crisis Situations

(1) Immediate or Short-Term Reactions to Crisis Situations

- Responds impulsively with inappropriate actions
- Freezes or becomes incapacitated
- Retreats from the situation
- Does not promptly or effectively communicate incident to other personnel who should take remedial action
- Displays a startled reaction or begins crying
- Places top priority on defending his/her innocence regarding the situation
- Shows signs of physiological reactivity such as trembling, sweating, dizziness, heart palpitations, shortness of breath, or fainting spells

(2) Reactions to Long-Term Effects of Accumulated Stress

- Exhibits deteriorating performance
- Develops mood changes
- Exhibits constant worrying
- Becomes hypersensitive to comments of others
- Complains of subjective feelings of tension
- Complains about pressures at work as well as home, family, financial status, etc.
- Exhibits decreased frustration tolerance
- Shows signs of developing substance dependency or abuse
- Appears chronically fatigued

- Calls in sick frequently
- Develops psychosomatic symptoms such as hypertension, gastric ulcers, migraine headaches, etc.

B. Behavioral Patterns Indicative of Generalized Emotional Instability

(1) Hostility Toward Authority

- Refuses to follow orders
- Exhibits screaming, obscenities, violence, arguments, or temper tantrums when questioned by superiors
- Shows arrogant and critical attitude toward company
- Violates standard operating procedures
- Does not follow appropriate chain of command
- Refuses to accept help from others
- Refuses to adhere to safety precautions
- Becomes easily agitated

(2) Illegal and Antisocial Behaviors

- Steals from organization
- Vandalizes facilities
- Engages in sabotage
- Provides inaccurate information to co-workers and superiors

(3) Irresponsibility

- Is careless in performing duties
- Is frequently tardy or absent
- Appears unconcerned with disciplinary measures
- Does not complete assignments
- Plays pranks on others on the job
- Chooses easiest or most apparent alternative

- Conducts personal business while on duty
- Acts impulsively

(4) Dependent Behavioral Patterns

- Is overly fearful of radiation exposure when no real danger is present
- Is unable to make own decisions and needs explicit instructions
- Shows signs of extreme timidity on the job
- Denies mistakes whenever possible, and makes excuses for proven mistakes
- Exhibits excessive need for approval

(5) Interpersonal Skill Deficiencies

- Shows lack of proper assertion
- Tends toward social isolation or withdrawal
- Is unable to effectively engage in casual or formal conversation
- Is unable to effectively transmit necessary information
- Stutters when trying to speak to others

(6) Deficiencies in Vigilance

- Displays low boredom tolerance
- Sleeps on the job
- Tries to create excitement
- Is inattentive to job duties

(7) Emotional and Thought Disturbances

- Shows no emotion at all
- Is overemotional (laughs, cries, becomes upset over minor things, etc.)
- Has insomnia
- Develops change in appetite

- Appears disoriented in time and space
- Is quite forgetful and has memory lapses
- Displays recurrent mood swings, from severe depression to extreme euphoria
- Displays lack of attention to personal appearance
- Exhibits excessive suspiciousness
- Expresses sensory hallucinations
- Displays difficulty in comprehending and responding to questions
- Creates and uses meaningless words or phrases
- Displays emotional responses which are inappropriate to the situation
- Exhibits delayed reaction time
- Displays decline in intellectual functioning

APPENDIX B

MEMBERS OF STANDARD INSTRUMENT CRITERIA DEVELOPMENT PANEL MEETING

APPENDIX B

MEMBERS OF STANDARD INSTRUMENT CRITERIA DEVELOPMENT PANEL MEETING

This Appendix lists the names, titles, and affiliations of the individuals who participated in our Standard Instrument Criteria Development Panel Meeting. This panel convened at Assessment Designs, Incorporated in Orlando, Florida, on August 22, 1980. The input collected at this meeting was incorporated into the standards for instrument adherence presented in Section 7.

The panel members in attendance included the following individuals:

- Cabot Jaffee, Ph.D., Co-Chairperson
  President, Assessment Designs, Inc.
- Fredric Frank, Ph.D., Co-Chairperson
  Vice-President, Assessment Designs, Inc.
- Robert Cohen, M.S.
  Professional Services Coordinator, Assessment Designs, Inc.
- Sandra Davis, Ph.D.
  Consulting Psychologist, Personnel Decisions, Inc.
- Sidney Lecker, M.D.
  President, Corporate Stress Control Services, Inc.
- Barbara Lindley, M.S.
  Professional Services Coordinator, Assessment Designs, Inc.
- David Lykken, Ph.D.
  Professor of Psychology and Psychiatry, University of Minnesota
- Richard McGee, Ph.D.
  Chief, Health Counseling and Rehabilitation Staff, Division of Medical Services, Tennessee Valley Authority
- Mike Rind, Esq.
  Senior Attorney, Legal Department, Baltimore Gas and Electric Co.
- Morris Roseman, Ph.D.
  Private Practice and Consulting Psychologist

APPENDIX C

RELEVANT RESEARCH STUDIES

APPENDIX C

RELEVANT RESEARCH STUDIES

1. INTRODUCTION

Previous research studies have investigated the personalities and performance of personnel employed in stressful occupations. These studies are particularly relevant to the selection of nuclear facility security personnel in that they are concerned with personnel employed in stressful positions. The following three sub-sections will review these research studies and will describe their relevance to nuclear facility security personnel selection.

2. AIR TRAFFIC CONTROLLERS

The position of air traffic controller (ATC), as with that of security personnel, has traditionally been considered a stressful occupation (Refs. 1, 2, 3, and 4). In addition, the two positions both require the analysis and reaction to information received from a control panel. ATC's receive information regarding the arrival and departure of specific aircrafts, while security guards receive information regarding the security of a nuclear power facility, e.g., the presence of intruders. The manner in which ATC and security personnel react, i.e., their resultant decisions and actions, can affect the safety of many individuals.

2.1 Physiology of Air Traffic Controllers

Dougherty (Ref. 5) reported the results of a two-year study on the physiological health of FAA employees, all of whom were located in the southwest region of the United States. The purpose of this study was to compare the health of air traffic controller specialists (ATCS) employees with that of non-ATCS employees and to compare the health of more experienced ATCS employees (journeymen) with that of non-ATCS employees. These results indicate that, as measured physiologically, journeymen ATCS employees and non-journeymen ATCS employees were experiencing more stress than non-ATCS employees. This is in agreement with previous studies which have shown significantly higher rates of stress-related symptoms in ATCS at certain experience levels and in comparison with non-ATCS employees (Refs. 1 and 2).

The relevance of this study to nuclear facility security personnel is that it shows that employees who are continuously involved in stressful situations show a higher rate of physiological abnormalities than employees who are not. In this respect, it should be noted that security personnel, unlike ATCS employees, may not experience continuous high levels of on-the-job stress. Intruders do not arrive at the

nuclear facility every hour on the hour. They may, however, experience continuous boredom and monotony, which in turn may be stressful.

2.2 Personality of Air Traffic Controllers

Karson and O'Dell (Ref. 3) examined the personality make-up, as measured by Form A of the Sixteen Personality Factor Questionnaire (16PF), of 11,047 persons employed as air traffic controllers (ATC) and 9,886 persons applying for that job. Comparisons were made between the ATC employees, the ATC applicants, and a sample of the general population (sample size = 1,127).

In general, the findings of this study were that the personalities of ATC employees and ATC applicants were generally similar to the general public. However, specific differences in the personality make-up of ATC employees and the general public did emerge.

The relevance of this study to nuclear facility security personnel is that slight differences did exist in the personality make-up of ATCs, traditionally considered a stressful occupation, and the general public.

2.3 Personality of Air Traffic Controllers and Its Relation to Job Performance

Karson (Ref. 6) conducted a research study investigating the relationship between the personality make-up of air traffic radar controllers, as measured by the 16PF, Form A, and the performance of the controllers, as measured by their peers and supervisors. The supervisors rated the controllers on elements such as judgment, cooperation, understanding, technical competence, and emotional stability. The sample of controllers consisted of 568 individuals employed at four air route traffic control centers.

The results of the study were:

1) higher performance ratings were significantly associated with lower warmth, lower anxiety, higher obsessive-compulsivity, and higher emotional sensitivity and ego strength scores;

2) the reliability of the criteria used was not demonstrated and it was concluded that they were subjective, unreliable, and invalid.

The relevance of this study to nuclear facility security personnel selection is that it reinforces the idea that reliable, valid criteria by which to validate the results of measurement instruments must be identified.

3. AIR FORCE PILOTS - PHYSIOLOGICAL STRESS IN SIMULATION

The job of jet pilot is a stressful one. Pilots must make instantaneous decisions during emergency situations. For this reason, the Air Force has been concerned with training pilots on how to react during stressful situations.

To accomplish this, the Air Force uses simulation training. That is, pilots are trained on a simulator that closely resembles the cockpit of an actual jet. The simulator and the surrounding environment are constructed in such a way that the trainees actually experience the total flight situation. The trainees sit in the simulated cockpit and perform flight maneuvers through simulated mission scenarios (e.g., a flight over mountainous terrain, a bombing mission). This, in turn, has evoked the question of whether simulation training can actually induce stress in pilot trainees. This question stems from the idea that the secure environment of the training simulator reduces the amount of stress associated with the training. To answer this question, the Air Force has conducted research measuring the amount of stress experienced by the trainees. One experimental study, Krahenbuhl (Ref. 7) found that pilot training was extremely stressful to trainees.

A study by Krahenbuhl (Ref. 8) was designed to investigate whether both a simulator and an aircraft can invoke a stress response when both devices are used to present the same mission scenario. The authors also investigated whether previous airborne flight experience altered one's stress response during simulation and whether previous simulation experience altered one's stress response during airborne flight. Stress was measured physiologically, that is, stress was assessed through a urinalysis, which is believed to be a quantifiable physiological expression of the general stress response as experienced by the individual (Ref. 9). The stressful situations were a power-on stall, spin prevention, and spin recovery. The subjects, 20 United States Air Force pilot training volunteers, experienced stressful situations in both the actual aircraft and the training simulator.

The relevance of this study to the selection of nuclear facility security personnel is that this study showed that physiological responses of stress can be induced through the use of a simulation. In commenting on this, Krahenbuhl, et al. (Ref. 8) stated that the stress response was due to the realism of the simulator and/or the scenario used in the experiment, since less realistic simulators have failed to evoke a stress response. Another relevant finding was that exposure to realistic simulation prior to actual experience alters the physiological stress response of the individual.

4. LAW ENFORCEMENT PERSONNEL

The relevance of investigating the research on law enforcement personnel, as related to nuclear facility personnel selection, is that research has shown that emotional stability is a crucial factor in determining the probability of success in law enforcement personnel (Ref. 10). Police are frequently engaged in stressful, oftentimes, life-threatening situations. The manner in which they respond to these situations could affect the safety of themselves and others.

A thorough review of the police selection literature conducted by Spielberger, et al. (Ref. 11) forms the basis for this review.

4.1 Biographical Data and Weighted Application Blanks

Traditionally, law enforcement agencies have set minimum height and weight requirements for selection of job applicants (Ref. 11). However, research which has examined the validity of these requirements for predicting success on the job have reported inconclusive and/or equivocal results (Refs. 12, 13, 14, 15, and 16). The courts have also contributed to the ambiguity associated with using height and weight as selection predictors. That is, cases such as Castio et al. v. Beecher et al., Hail v. White, and Smith v. Trogran have upheld the use of minimum size standards, whereas Hardy v. Stampf and Smith v. City of East Cleveland have found such hiring practices to be unjustifiable (Ref. 17). Spielberger, et al. (Ref. 11) reviewed the research on the validity of selecting police personnel using biographical variables (Refs. 18, 19, 20, and 21), and concluded that of all the predictors investigated only previous military experience appeared to predict the performance of police officers in a reasonably consistent manner.

4.2 Personality Measures

Research studies on police selection have been conducted which investigated the use of personality measures such as the Minnesota Multiphasic Personality Inventory (MMPI) (Refs. 17, 22, 23, 24, 25, and 26),the California Psychological Inventory (CPI) (Ref. 27), and Eysenck Personality Inventory (EPI) (Ref. 28). Spielberger, et al. (Ref. 11) stated that in most of the concurrent criterion-related validity studies in which the MMPI, CPI, and EPI were examined, the personality profiles of successful police officers were not different from those of the general population.

4.3 Situational Simulations

Spielberger, et al. (Ref. 11) stated that the research studies on the validity and contribution of situational simulations to the prediction of job-related behaviors of police officers (Refs. 30, 31, 32, 33, 34, and 35) have shown promising results.

4.4 Law Enforcement Research: Summary

The relevance of the law enforcement research studies to nuclear facility security personnel selection is indirect in that these studies did not investigate the selection of police officers based on their emotional stability, per se.

In examining whether a selection instrument is a valid predictor, consideration must also be given to whether the criteria being used are valid. In this respect, the police officer studies must be investigated to discover whether the criteria used to differentiate successful from unsuccessful performance are acceptable. Kent and Eisenberg (Ref. 36) reviewed ten years of research on police selection and concluded that a basic reason that a useful and valid selection procedure has not yet been developed is that good criterion measures have not been identified.

5. RELEVANT RESEARCH - SUMMARY

The most striking conclusion to be reached regarding the relevance of this previous research to the selection of nuclear facility security personnel is that few of the above studies were specifically concerned with predicting the emotional stability, or the stress tolerance of the participants. Two reasons for this are:

1) Researchers have not been able to develop instruments which are reliable and valid predictors of emotional stability, as it relates to on-the-job performance.

2) Even more critical is that criterion measures of emotional instability on the job have not been identified. That is, criteria for measuring emotionally unstable performance have not yet been adequately developed.

APPENDIX C

REFERENCES

1. J.D. Dougherty, D.K. Trites, and R. Dille, "Self-reported Stress-related Symptoms Among Air Traffic Control Specialists (ATCS) and Non-ATCS Personnel," Aerospace Medicine 36, 956 (1965). Available in public technical libraries.

2. G.W. Hauty, D.K. Trites, and W. Berkley, "Frequency of Shift Rotation at Air Traffic Control Facilities and Incidence of Stress-related Symptoms," Aerospace Medicine 36, 350 (1965). Available in public technical libraries.

3. S. Karson and J.W. O'Dell, "Personality Make-up of the American Air Traffic Controller," Aerospace Medicine 45, 1001-1007 (1974). Available in public technical libraries.

4. P. Lehman, "Job Stress: Hidden Hazard," Job Safety and Health 2, 4-10 (1974). Available in public technical libraries.

5. J.D. Dougherty, "Cardiovascular Findings In Air Traffic Controllers," Aerospace Medicine 38, 26-30 (1967). Available in public technical libraries.

6. S. Karson, "Some Relations Between Personality Factors and Job Performance Ratings in Radar Controllers," Aerospace Medicine 40 (8), 823-826 (1969). Available in public technical libraries.

7. G.S. Krahenbuhl, J.R. Marett, and G.B. Reid, "Stress and Simulation in Pilot Training," AFHRL-TR-78-95, Williams Air Force Base, A2: Flying Training Division, Air Force Human Resources Laboratory, February 1979. Available for purchase from National Technical Information Service, Springfield, VA 22161.

8. G.S. Krahenbuhl, J.R. Marett, and N.W. King, "Catecholamine Excretion In T-37 Flight Training," Aviation, Space, and Environmental Medicine 48, 405-408 (1977). Available in public technical libraries.

9. U.S. Euler, "Quantification of Stress by Catecholamine Analysis," Clinical Pharmacology and Therapeutics 5, 398-408 (1964). Available in public technical libraries.

10. M.E. Baehr, D.R. Saunders, E.C. Froemel, and J.E. Furcon, "The Prediction of Performance for Black and for White Police Patrolmen," Professional Psychology 2, 46-57 (1971). Available in public technical libraries.

11. C.D. Spielberger, J.C. Ward, and H.C. Spaulding, "A Model for the Selection of Law Enforcement Officers," in Police Selection and Evaluation: Issues and Techniques, C.D. Spielberger, Ed. (Hemisphere Publishing Company, Washington, D.C., 1979).

12. C.A. Dempsey, "A Study of Police Height Requirements," The Police Chief 41, 34-35 (1974). Available in public technical libraries.

13. L.S. Goldstein, "Comment on ("A Question of Height" by R.I. Hoobler and J.A. McQueeney in The Police Chief, 1973)," The Police Chief 41, 11 (1974). Available in public technical libraries.

14. B.E. Halling, "Comment on ("A Question of Height" by R.I. Hoobler and J.A. McQueeney in The Police Chief, 1973)," The Police Chief 41, 10-11 (1974). Available in public technical libraries.

15. W.B. Kollender and J.A. McQueeney, "A Question of Height - Additional Thoughts," The Police Chief 44, 56-58 (1977). Available in public technical libraries.

16. B. Prelutsky, "Second Thoughts," Human Behavior 3, 9-10 (1974). Available in public technical libraries.

17. A.E. Shealey, "Height and Weight as Factors in the Selection of Police Officers," in C.D. Spielberger, Ed. Proceedings of the National Workshop on the Selection of Law Enforcement Officers, Human Resources Institute, Monograph Series Three: No. 3, College of Social and Behavioral Sciences, University of South Florida, Oct. 1979. Available for purchase from the National Technical Information Service, Springfield, VA 22161.

18. S.P. Azen, H.R. Montgomery, H.M. Snibbe, J.M. Fabricatore, and H.H. Earle, "Predictors of Resignation and Performance in Law Enforcement Officers," American Journal of Community Psychology 2, 79-86 (1974). Available in public technical libraries.

19. S.P. Azen, H.M. Snibbe, and H.R. Montgomery, "A Longitudinal Study of Success and Performance in Law Enforcement Officers," American Journal of Community Psychology 2, 79-86 (1974). Available in public technical libraries.

20. L.S. Goldstein, "Characteristics of Police Applicants," The Police Chief 39, 58-60 (1972). Available in public technical libraries.

21. R.J. Levy, "Predicting Police Failures," Journal of Criminal Law, Criminology, and Police Science 58, 265-276 (1967). Available in public technical libraries.

22. J.A. McAllister, "A Study of the Prediction and Measurement of Police Performance," Police 15, 58-64 (1970). Available in public technical libraries.

23. B. Barnabas, "Profile of a Good Police Officer," Law and Order 24, 32-44 (1976). Available in public technical libraries.

24. J. Gottesman, The Utility of the MMPI in Assessing the Personality Patterns of Urban Police Applicants, Stevens Institute of Technology, Hoboken, N.J., 1975.

25. J.F. Hooke and H.H. Krauss, "Personality Characteristics of Successful Police Sergeant Candidates," Journal of Criminal Law, Criminology, and Police Science 62, 104-106 (1971). Available in public technical libraries.

26. A.E. Shealey, "The MMPI and the Meyer-Briggs Type Indicator in Police Selection," pp. 95-102 in C.D. Spielberger and H.C. Spaulding, Eds. Proceedings of the National Working Conference on the Selection of Law Enforcement Officers, 1977. Available for purchase from the National Technical Information Service, Springfield, VA 22161.

27. E.J. Sullivan and R.D. Cloward, "Psychological Testing for Police Selection," RIGJC Report 72-2101A, University of Rhode Island, 1976. Available for purchase from National Technical Information Service, Springfield, VA 22161.

28. R. Hogan, "Personality Characteristics of Highly Rated Policemen," Personnel Psychology 24, 679-686 (1971). Available in public technical libraries.

29. L.C. Parker and M.C. Roth, "The Relationship Between Self-disclosure, Personality, and a Dimension of Job Performance of Policemen," Journal of Police Science and Administration 1, 282-287 (1973). Available in public technical libraries.

30. C.A. Fenster and B. Locke, "Neuroticism Among Policemen: An Examination of Police Personality," Journal of Applied Psychology 57, 358-359 (1973). Available in public technical libraries.

31. P.F. D'Arcy, in "New York City Assessment Center Program Helps Test Managerial Competence," The Police Chief 41, 52 (1974). Available in public technical libraries.

32. M.D. Dunnette and S.J. Motowidlo, "Police Selection and Career Assessment," (Grant No. 73-NI-99-0018-G and 74-NI-99-0001-G). Washington D.C.: National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice, November 1976. Available for purchase from National Technical Information Service, Springfield, VA 22161.

33. J.F. Gavin and J.W. Hamilton, "Selecting Police Using Assessment Center Methodology," Journal of Police Science and Administration 3, 166-176 (1975). Available in public technical libraries.

34. D.A. Kent, C.R. Wall, and R.L. Bailey, "A New Approach to Police Personnel Decisions," The Police Chief 41, 72-77 (1974). Available in public technical libraries.

35. R.B. Mills, "Simulated Stress in Police Recruit Selection," Journal of Police Science and Administration 4, 179-186 (1976). Available in public technical libraries.

36. D.A. Kent and T. Eisenberg, "The Selection and Promotion of Police Officers: A Selected Review of the Literature," The Police Chief 39, 20-29 (1972). Available in public technical libraries.

APPENDIX D

DEFINITION OF TERMS

APPENDIX D

DEFINITION OF TERMS

1. CONCURRENT VALIDITY - a type of criterion-related validity. Concurrent validity is computed by correlating results on the predictor with results on the criterion for a group of individuals who are presently on the job. These criterion measures are collected at the same time or so close in time such that one can not assume that changes in conditions or subjects have occurred (Ref. 1).

2. CONTENT VALIDITY - a qualitative statement based on the judgments of various experts, regarding the similarity between the content of the predictor and the content of the position being tested for.

3. PREDICTIVE VALIDITY - a type of criterion-related validity. Persons' measures on a predictor are correlated with the same persons' measures on the criterion, but a period of time elapses between the collection of these two measures. During this period of time some identifiable, interviewing event occurs, such as the persons acquiring experience on the job (Ref. 1).

4. VALIDITY - the degree to which a predictor measures what it is intended to measure or achieves its purpose.

5. RELIABILITY - consistent or dependable; repeatable; refers to the accuracy or consistency of measurement and to whether a set of measures obtained can be replicated. It is quantitatively computed using correlation (Ref. 2).

6. TEST-RETEST RELIABILITY - a correlation between persons' scores on a test administered at one point in time and those same persons' scores on the same test administered at another point in time.

7. EQUIVALENT FORMS RELIABILITY - a correlation between persons' scores on one measurement instrument and those same persons' scores on a similar, but different instrument.

8. INTERNAL-CONSISTENCY RELIABILITY - a correlation between two parts of the same measurement instrument. This is accomplished by dividing one instrument into two or more equivalent parts.

9. INTER-RATER RELIABILITY - a measure of the extent to which two or more people evaluating the same behaviors are in agreement with one another. It is computed by correlating the evaluation scores of various evaluators.

10. STANDARD ERROR OF MEASUREMENT - The standard error of measurement, which is derived from the reliability, is used to provide an estimate of how accurate an individual's test score is. With a low standard error of measurement, considerable confidence can be placed in the accuracy of a test score, and the interval necessary for spanning the actual score will be quite small (Ref. 3). This interval, or range, is actually a zone of inaccuracy.

11. CRITERION-RELATED VALIDITY - A quantitative method of measuring how well performance on a measurement instrument predicts performance on the job. It is computed by correlating persons' scores on a predictor with their scores on a criterion measure.

12. CONSTRUCT VALIDITY - a measure of whether the predictor actually measures the constructs it is purported to measure. With construct validity, the unobservable constructs measured by the predictor must be inferred from observed behavior. Construct validity is qualitatively determined by obtaining the judgments of experts and quantitatively determined by computing convergent and discriminant validities (Ref. 3).

13. CONVERGENT VALIDITY - a quantitative method of computing construct validity. It is computed by correlating the results of various predictors which measure the same construct. A high correlation indicates high construct validity (Ref. 4).

14. DISCRIMINANT VALIDITY - a quantitative method of computing construct validity. It is computed by correlating the results of predictors which measure different constructs. A low correlation indicates high construct validity (Ref. 4).

15. PREDICTOR - a measurement instrument which is being used to predict a person's future behavior or performance.

16. CRITERION - a measure of the behavior or performance that one is attempting to predict (e.g., performance on the job).

17. PREDICTIVE VALIDITY - a type of criterion-related validity, in which a period of time elapses between the collection of measures on the measurement instrument and on the criterion. During this period of time something discernible happens, such as training or the acquisition of experience (Ref. 1).

18. BASE RATE - the proportion of employees currently employed who are successful.

19. SELECTION RATIO - the proportion of persons selected from those who apply.

20. COST OF AN ERRONEOUS ACCEPTANCE - the cost associated with selecting an applicant for a position who subsequently proves to be unsuccessful. This includes such costs as recruiting, training, and salary paid to the employee.

21. COST OF AN ERRONEOUS REJECTION - the cost associated with refusing employment to an applicant who would have been a successful employee.

22. COST OF THE MEASUREMENT INSTRUMENT - the total cost of developing, administering, and maintaining the instrument.

APPENDIX D

REFERENCES

1. R.M. Guion, "Recruiting, Selection, Job Placement," in the Handbook of Industrial and Organizational Psychology, M.D. Dunnette, Ed. (Rand McNally College Publishing Company, Chicago, 1976), Chap. 18, pp. 777-828.

2. American Psychological Association, Standards for Educational and Psychological Tests, Washington D.C., 1974. Available from APA.

3. E. Lemke and W. Wiersma, Principles of Psychological Measurement, Rand McNally College Publishing Company, Chicago, 1976.

4. D.T. Campbell and D.W. Fiske, "Convergent and Discriminant Validation by the Multitrait-multimethod Matrix," Psychological Bulletin 56(2), 81-105 (1959). Available in public technical libraries.

APPENDIX E

SAMPLE FORM FOR DOCUMENTATION OF NUCLEAR PERSONNEL SELECTION PROCEDURE RESULTS

APPLICANT SELECTION PRODCEDURE RESULTS

I. Introduction

The purpose of this form is for documentation of selection procedure results obtained by the prof ssional administrator/evaluator. Its sole purpose is to forward relevant information regarding on-the-job indices of emotional instability back to the individual responsible for the hiring process pertaining to the applicant herein.

It is stressed that this information is confidential in nature. Prior to administration of any selection instruments, the applicant must sign an informed consent form indicating that he/she agrees to have the results of the evaluation process forwarded back to the hiring individuals. Subsequent to this, it becomes the responsibility of the hiring personnel to maintain these data in confidential files. If release of information to other individuals is desired at a later date, the applicant must sign a new consent form which lists the new recipients.

The attached information indicates that the applicant has signed an informed consent form for release of interpreted data to hiring personnel. This consent form, as well as test materials, are to be retained by the professional.

II. Information Regarding Position Desired by Applicant

A. Name of Organization: _____

B. Specific Title of Position: _____

1. Section/Unit: _____

2. Department: _____

3. Degree of Access to Vital or Protected Areas Associated with the Position: _____

4. Critical Job Responsibilities (List or attach copy of job description): _____

C. Is Applicant Currently Employed in this Facility?
Yes: _____ No: _____

If "yes", indicate current position: _____

III. Instruments Administered

In accordance with recommendations relating to the specific position in question, the following instruments were administered to the applicant in the sequence indicated:

	Instrument	Administrator	Date of Administration
1)	_____	_____	_____
2)	_____	_____	_____
3)	_____	_____	_____
4)	_____	_____	_____

IV. Combined Results of Selection Instruments

A) Instructions: The results to be presented are phrased in behavioral terms that are predictive of overt actions indicative of emotional instability. Only those behaviors detected which might pose a risk to the safe operation of the nuclear facility are included here. These include behavioral tendencies indicative of overall risks, as well as relevant behaviors with regard to the Critical Job Responsibilities previously outlined for this specific position.

B) Specific Results (Attach extra pages as needed) _____

V. Recommendations to Hiring Personnel

Based on the preceding Results section, and in accordance with the nature of the specific demands of the position for which application was made, the following decision is recommended based on the relative emotional stability of the applicant:

1) Unconditional Acceptance _____

2) Conditional Acceptance* _____

3) Rejection (with opportunity for reevalaution at a later date). _____

4) Evaluation by another professional** _____

(Date of Recommendations) _____ (Signature)

Full name of professional examiner including any pertinent degrees or certifications.

* If an applicant is recommended for conditional acceptance, this means that the individual appears to be generally emotionally stable, however the applicant should be closely observed for specific behavioral tendencies noted (outlined on an attached page) which might pose risks and could not be definitively detected by this evaluation.

**The professional examiner recommends evaluation by a second professional due to inconclusive results of the initial evaluation.

APPENDIX F

SAMPLE FORMS FOR DOCUMENTATION OF NRC AND USER ORGANIZATIONS FOR IMPLEMENTATION OF STANDARDS

FORM A: INITIAL DOCUMENTATION OF ORGANIZATIONAL IMPLEMENTATION OF STANDARDS

_____ _____

Name of Organization Date of Submission to NRC

_____

Position Title to Be Evaluated

The purpose of this form is to provide the Nuclear Regulatory Commission with written documentation of the above named organization's intent to comply with the NRC's recommended Standards and Selection Procedures for Nuclear Facility Personnel Selection. Thus, this organization will use combinations of the following measurement instruments in accordance with the recommendations made by the NRC for the designated positions (please place a checkmark next to instruments to be used).

1. Minnesota Multiphasic Personality Inventory (MMPI) _____
2. Sixteen Personality Factor Questionnaire (16PF) _____
3. The Clinical Interview _____
4. Situational Simulations _____

- This organization plans to utilize internal_____ external_____ (check one) professionals to administer and interpret instrument results
- This organization plans to implement these selection procedures by the following date:_____

Any possible exceptions to the recommended procedures which pertain to specific positions, as well as any questions/problems regarding implementation, are indicated on the attached page(s).

_____ _____

Date Signed (Signature of Authorized Representative)

Full Name

Position within the organization

FORM B: REQUEST FOR REVIEW OF ALTERNATIVE SELECTION PROCEDURES DESIRED FOR USE BY ORGANIZATION

_____ _____

Name of Organization Date of Submission to NRC

_____

Position Title(s) to Be Evaluated

The purpose of this form is to document adherence to standards for selection instruments prescribed by the Nuclear Regulatory Commission, utilizing specific instruments different than those recommended by the Nuclear Regulatory Commission.

I. The selection procedures desired for use by this organization include administration of the following instruments:

(1) _____

(2) _____

(3) _____

(4) _____

(5) _____

- If the procedures desired for use are planned to vary according to position, pages should be attached indicating which procedures apply to which positions (or categories of positions) along with accompanying justification for these variations.

- For each instrument listed above, a separate justification statement (including relevant references) applying to each of the following standards should be included on this form (see Part II). Attach extra pages as necessary.

II. Instrument: _____

Manufacturer: _____

Recommended NRC Instrument for Which this is a Substitution: _____

(1) Reliability:

(2) Validity:

(3) Compliance with Legal Issues, Labor Relations, and Uniform Guidelines on Employee Selection procedures:

(4) Personal Effects on Applicants:

(5) Susceptibility to Faking

(6) Reevaluation Considerations

III. It is understood by the user organization and instrument manufacurer(s) that the use of the instruments advocated here must conform to the standards which are universally applica-

ble to all selection instruments, as well as the selection procedures as a whole.*

The user organization, upon submitting this request, will wait for a written reply from the NRC regarding the acceptability of the proposed instrument(s). It is understood that this reply may indicate any areas of deficiency where further justification/ documentation is necessary.

(Signature of Authorized Representative)

Full Name

Position within the organization

Date Signed

*Refer to Section 7.3 of this document for the discussion of these standards.

NRC FORM 335 (7-77)

U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

1. REPORT NUMBER (Assigned by DDC)
NUREG/CR-2075

4. TITLE AND SUBTITLE (Add Volume No., if appropriate)
Standards for Psychological Assessment of Nuclear Facility Personnel

2. (Leave blank)

3. RECIPIENT'S ACCESSION NO.

7. AUTHOR(S)
Fredric D. Frank, Barbara S. Lindley, and Robert A. Cohen

5. DATE REPORT COMPLETED
MONTH: December YEAR: 1980

9. PERFORMING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code)
Assessment Designs, Incorporated
ADI Court - 601 North Ferncreek
Orlando, FL 32803

DATE REPORT ISSUED
MONTH: July YEAR: 1981

6. (Leave blank)

8. (Leave blank)

12. SPONSORING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code)
U. S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Division of Facility Operations
Washington, DC 20555

10. PROJECT/TASK/WORK UNIT NO.

11. CONTRACT NO.
FIN B1083

13. TYPE OF REPORT
Technical Report

PERIOD COVERED (Inclusive dates)

15. SUPPLEMENTARY NOTES

14. (Leave blank)

16. ABSTRACT (200 words or less)

The subject of this study was the development of standards for the assessment of emotional instability in applicants for nuclear facility positions. The investigation covered all positions associated with a nuclear facility. Key findings were that emotional instability is a multi-dimensional concept; no single instrument by itself is capable of measuring emotional instability; few studies have been conducted in a nuclear setting aimed at determining the predictive validity of various selection instruments with respect to emotional stability; and standard criteria for evaluating instruments require careful consideration of psychometric principles and legal considerations. Conclusions reached in this investigation focused on the ingredients of an integrated selection system including the use of personality tests, situational simulations, and the clinical interview; the need for professional standards to ensure quality control; the need for a uniform selection system as organizations vary considerably in terms of instruments presently used; and the need for an on-the-job behavioral observation program.

17. KEY WORDS AND DOCUMENT ANALYSIS

17a. DESCRIPTORS

17b. IDENTIFIERS/OPEN-ENDED TERMS

18. AVAILABILITY STATEMENT
Unlimited

19. SECURITY CLASS (This report)
Unclassified

21. NO. OF PAGES

20. SECURITY CLASS (This page)
Unclassified

22. PRICE
$

NRC FORM 335 (7-77)

UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, D. C. 20555

July 31, 1981

MEMORANDUM FOR: Distribution List

FROM: Frank P. Gillespie, Chief
Safeguards Branch
Office of Nuclear Regulatory Research

SUBJECT: NUREG/CR-2075 - STANDARDS FOR PSYCHOLOGICAL ASSESSMENT OF NUCLEAR FACILITY PERSONNEL

The Office of Nuclear Regulatory Research has recently published the subject document. Information from this report has been included in a proposed Regulatory Guide in support of the proposed rule "Power Reactor Access Authorization Rule." In order to assist us at an early stage in assessing the need for a Guide and provide for the earliest possible evaluation of comments, you have been supplied with a copy of the report. If you have any questions, comments, or suggestions after you review the report, please forward them to this Office for consideration.

Any suggestions or opinions you have concerning the need for additional guidance in the application of Safeguards requirements would also be welcome.

Frank P. Gillespie, Chief
Safeguards Branch
Office of Nuclear Regulatory Research

Enclosure:
NUREG/CR-2075