AFRL-RH-WP-TR-2020-0103

Air Force Personnel Center Best Practices Guide:

Legal and Ethical Issues in Personnel Testing

Art Gutman

PDRI, an SHL Company

November 2020

Interim Report

AIR FORCE RESEARCH LABORATORY
711TH HUMAN PERFORMANCE WING,
AIRMAN SYSTEMS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th Air Base Wing Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RH-WP-TR-2020-0103 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

CARRETTA.THOMAS.R.1228984929 Digitally signed by CARRETTA.THOMAS.R.1228984929 Date: 2020.12.14 13:35:32 -05'00'

THOMAS R. CARRETTA
Work Unit Manager
Performance Optimization Branch
Airman Biosciences Division

WILLIAMS.LOGAN.ANDREW.1273597634 Digitally signed by WILLIAMS.LOGAN.ANDREW.1273597634 Date: 2020.12.14 14:47:58 -05'00'

LOGAN A. WILLIAMS
Airman Readiness Optimization CRA
Performance Optimization Branch
Airman Biosciences Division

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
20-11-20	Interim	March 2019 – November 2020

4. TITLE AND SUBTITLE

Air Force Personnel Center Best Practices Guide: Legal and Ethical Issues in Personnel Testing

5a. CONTRACT NUMBER
FA8650-14-D-6500, Task Order 0007

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
Art Gutman

5d. PROJECT NUMBER
5329

5e. TASK NUMBER
09

5f. WORK UNIT NUMBE
H0SA (532909TC)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
PDRI, an SHL Company
1911 N. Fort Myer Drive
Suite 410
Arlington, VA 22209

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Material Command
Air Force Research Laboratory
711th Human Performance Wing
Airman Systems Directorate
Airman Biosciences Division
Performance Optimization Branch
Wright-Patterson AFB, OH 45433

Air Force Personnel Center
Strategic Research and Analysis Branch
550 C St West, Ste. 45
JBSA-Randolph, TX 78150-4747

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
711 HPW/RHBC

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RH-WP-TR-2020-0103

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A: Approved for public release. MSC/PA-2020-0277 AFRL-2020-0452, cleared 9 December 2020

13. SUPPLEMENTARY NOTES
Report contains color.

14. ABSTRACT

This series of reports consolidates the experience, wisdom, and tools the Air Force has accumulated in its selection and classification work, and blends them with best practice recommendations from industry. This entry addresses recommendations and best practices for understanding and evaluating critical legal issues relating to adverse impact. The chapter is divided into six sections including: A comparison of adverse impact to other forms of discrimination; an overview of major court rulings related to adverse impact; a discussion of major issues relating to the Uniform Guidelines on Personnel Selection; an examination of how to create and validate test batteries; how and when to employ content validity strategies; and a review of court rulings related to minimizing adverse impact.

15. SUBJECT TERMS
Selection, Validation, Discrimination, Adverse Impact, Court Rulings, Uniform Guidelines

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	b. ABSTRACT	c. THIS PAGE			Thomas R. Carretta
Unclassified	Unclassified	Unclassified	SAR	49	19b. TELEPHONE NUMBER *(Include Area Code)* N/A

TABLE OF CONTENTS

FOREWORD iii
EXECUTIVE SUMMARY iv
Background/History 1
Air Force Human Resources Laboratory 1
The Rise of the Strategic Research and Assessment Branch 1
AFPC/DSYX Program Overview 2
AFPC/DSYX Organizational Structure 2
Synergistic Relationships 2
The AFPC/DSYX Contribution to Human Capital Management and Strategic Human Resources Management through Mission Alignment 3
The DSYX Testing Toolbox 3
General Ability/Aptitude Tests 3
Vocational Interests 4
Personality 4
Miscellaneous/Specialty 5
The DSYX Expertise and Resources Toolbox 6
Forward Looking: The Future of AFPC/DSYX 7

Increased Effort to have AFPC/DSYX Expertise, Services, and Interventions Recognized throughout the Air Force 7

Improved Technology 7
Improved Access to Data 7
Exiting the Operational Testing Domain 7
Repeatable and Scalable Processes 8
1.0 ETHICAL AND LEGAL CONSIDERATIONS IN PERSONNEL TESTING 9
1.1 Introduction 9
1.2 Section 1: Overview of Types of Discrimination 9
1.2.1. Facial Discrimination 9
1.2.2. Disparate Treatment 10
1.2.3. Adverse Impact 11
1.2.4. Pattern or Practice 12
1.3 Section 2: Landmark Adverse Impact Rulings 13
1.3.1. Griggs v. Duke Power (1971) 14
1.3.2. Albemarle v. Moody (1975) 15
1.3.3. Washington v. Davis (1976) 15
1.3.4. Dothard v. Rawlinson (1977) 16
1.3.5. Connecticut v. Teal (1982) 16

1.3.6. Watson v. Fort Worth Bank (1988) 16
1.3.7. Wards Cove v. Atonio (1989) 17
1.4 Section 3: Uniform Guidelines on Personnel Selection Procedures 18
1.4.1. Role of the Psychological Profession 19
1.4.2. Content Validity 20
1.4.3. Multiple Hurdle and Multiple Selection Procedures 21
1.4.4. The Four-Fifths (Or 80% Rule) 22
1.5 Section 4: A Best Practice Procedure for Job Analysis and Developing and Validating a Text Battery 24
1.5.1. Step 1: Comprehensive Job Analysis 25
1.5.2. Step 2: Development of a Test Battery 26
1.5.3. Step 3: Development of the EKT Test 28
1.5.4. Step 4: Criterion Validation of the Test Battery 28
1.6 Section 5: How and when to employ content validity strategies 30
1.7 Section 6: Court rulings related to minimizing adverse impact 31
1.7.1. Method 1: Subgroup Norming 32
1.7.2. Method 2: Banding 32
1.7.3. Method 3: Alternative Tests or Combinations 33
1.7.4. Method 4: Manipulating Test Content 34
1.7.5. Method 5: Discarding Test Results 35
2.0 CONCLUSIONS 37
REFERENCES 38

LIST OF TABLES

Table 1. The McDonnel Douglas Scenario 11
Table 2. The Adverse Impact Scenario 12
Table 3. Supreme Court Landmark Rulings 14
Table 4. Justice O'Connor's Proposed Provisions 17
Table 5. The Flip Flop Rule 23
Table 6. Final Test Battery Plan 27
Table 7. Methods for Reducing Adverse Impact 32

FOREWORD

This report is one of a series of that compile the best of the experience, wisdom, and tools that the Air Force has accumulated in its selection and classification work, and best practices from industry and academia. These reports draw upon the experiences of the Air Force Personnel Center, Strategic Research and Assessment (AFPC/DSYX) branch and leading researchers and practitioners in the field of Industrial/Organizational Psychology to provide guides to cover a variety of topics. Each begins with a chapter describing AFPC/DSYX and the background of their research to provide context for the series. This report addresses major issues relating to legal and ethical considerations in testing, with an emphasis on adverse impact and associated legal issues.

EXECUTIVE SUMMARY

This series of reports is intended to consolidate the experience, wisdom, and tools that the Air Force has accumulated in its selection and classification work, and to blend these with best practice recommendations from industry. The reports cover a wide variety of material, including chapters on test development and validation, selection/classification model development, reporting/briefing results, and ethical and legal considerations. The goal is to ensure consistency as the Air Force Personnel Center Strategic Research and Assessment Branch (AFPC/DSYX) continues to develop assessments and refine selection and classification models for a large number of Air Force career fields.

We begin with an introduction to AFPC/DSYX. The background and history are covered, describing how the Air Force Human Resources Laboratory and its elimination left a need for providing research in human capital management. That was resolved in 2010 with funding to create DSYX which is intended to review, evaluate, develop, validate, and manage personnel programs to improve recruiting, selection, classification, and utilization of military personnel. The chapter describes how AFPC/DSYX contributes to strategic human capital management, tools it makes available for testing, experience and expertise it provides, and looks ahead to the future and how AFPC/DSYX can build on its capabilities.

The body of this report provides recommendations and best practices for understanding and evaluating critical legal issues relating to adverse impact. The chapter is divided into six sections including: A comparison of adverse impact to other forms of discrimination; an overview of major court rulings related to adverse impact; a discussion of major issues relating to the *Uniform Guidelines on Personnel Selection*; an examination of how to create and validate test batteries; how and when to employ content validity strategies; and a review of court rulings related to minimizing adverse impact.

Introduction to the Air Force Personnel Center, Strategic Research and Assessment Branch

Background/History

Human Capital Management Mandates. The Air Force Policy Directive, AFPD 36-XX, Air Force Personnel Assessment Program, raised the bar for validation of Air Force operations affecting human capital management. The policy directive laid out Air Staff-defined objectives in support of both 1) Department of Defense (DoD) initiatives, such as the Testing Modernization Program, supported by major influxes of research and development funding and 2) the Human Capital Annex of the Air Force Strategic Personnel Plan (moving ahead with several active Air Force-level working groups). The Air Force's way forward in support of these flow-down mandates included both the objectives and the scope of this initiative:

- Establish processes to apply scientific analysis and technology in support of recognized best practices to support personnel assessment. The goal of the Air Force Personnel Assessment Program is to support effective force management by ensuring that the right persons having the right aptitudes, characteristics, skills, and abilities are identified and accessed into the Air Force, are properly trained, and then optimally utilized to support the Air Force mission.
- The Air Force Personnel Assessment Program includes, but is not limited to, selection and classification, promotion, and proficiency assessment; and survey capability for assessing attitudes and opinions, job performance, and Air Force Specialty (AFS) requirements and characteristics.

Air Force Human Resources Laboratory

In 1968, the broad personnel research efforts (e.g., manpower, personnel, training) from various programs across the Air Force were consolidated into the Air Force Human Research Laboratory (AFHRL). The name "Air Force Human Resources Laboratory" was only used as the official designation for the combined program from 1968 to 1991. However, it was the name used for the longest period of time and is the one that has the greatest familiarity to professionals, in and out of the government, with an interest in military psychology. The antecedents of AFHRL can be traced to the Psychological Research Units of the Aviation Psychology Program in the Army Air Corps during World War II. After the Air Force became a separate service in 1947, AFHRL was called the Human Resources Research Center (1949-1953), Personnel and Training Center (1954-1958), Personnel Laboratory (1958-1962), and Personnel Research Laboratory (1962-1968). In 1991, the name Air Force Human Resources Laboratory was retired and the mission was absorbed by successor organizational units within the Armstrong Laboratory (1991-1996) and the Air Force Research Laboratory (1997-1999). In 1999, the personnel research function in the Air Force (Manpower and Personnel Research Division) was eliminated, leaving no organizational entity for research in the domains of personnel selection and classification.

The Rise of the Strategic Research and Assessment Branch

The need for research in strategic human capital management within the Air Force did not end with the elimination of AFHRL funding. After the elimination of AFHRL, minimal funding was

provided to manage testing-related contracts and provide basic support for operational testing programs. In 2010, additional funding was provided to create the AFPC/DSYX program and several billets were created to continue the work that ended with the elimination of AFHRL in 1999.

AFPC/DSYX Program Overview

With the additional funding, the AFPC/DSYX program was tasked to review, evaluate, develop, validate, and manage personnel programs to improve recruiting, selection, classification, and utilization of military personnel. The current responsibilities of AFPC/DSYX include Air Force- and Department of Defense-related testing programs, research and analysis, and development and validation of new assessment processes and measures. The AFPC/DSYX program now develops person-job match screening processes to support optimal personnel utilization for the entire personnel life cycle including pre-recruiter job exploration (e.g., interest inventories, realistic job previews); applicant assessment, screening, and classification of recruits (e.g., cognitive, personality, psychomotor, occupation-specific assessment of skills), retraining, and specialized assignments.

The DSYX program also helps maintain a mission-ready force by managing Air Force Specialty Code (AFSC) structures using scientific standards to establish desirable and mandatory occupational entry requirements and adjust occupational structures to optimize training investment, career progression, utilization, and retention for total force integration. Thus, the ultimate purpose of the AFPC/DSYX program is to provide: 1) consultation to program managers and Air Force leadership on selection and classification issues, 2) development, revision, and validation of personnel tests, 3) technical oversight of the operational testing program, and 4) management of contracts in support of personnel-related research.

AFPC/DSYX Organizational Structure

The DSYX branch is now embedded within the AFPC Directorate of Staff. As previously mentioned, while no longer supported by a multitude of scientists and psychologists, AFPC/DSYX provides an array of services and tools similar to AFHRL. The current structure of AFPC.DSYX includes the branch chief, a program manager, seven personnel research psychologists, and two research assistants. While many of the tasks assigned to DSYX and much of the funding to accomplish them come from Air Staff (A1) and Air Force Testing Policy (AF/A1PT), AFPC/DSYX is officially under the command of AFPC.

Synergistic Relationships

The Air Force Personnel Center Promotions, Evaluations, and Recognition Branch (AFPC/DP3SP) manages the operational personnel testing program. Thus, while AFPC/DSYX has the responsibility of developing and validating the tests within the personnel testing program, the operational responsibility of military testing resides with AFPC/DP3SP. The one current exception is the Pilot Candidate Selection Method (PCSM; described later in this report) which has been developed, validated, and operationally maintained by AFPC/DSYX.

The Air Force Recruiting Service (AFRS) Operations Division's Analysis Branch (AFRS/RSOA) supports DSYX through participation in the regular working group conference calls with AF/A1PT and DSYX, pre-accession process advisories, data collection facilitation, collaborative ad hoc analyses, and unrestricted access to relevant operational data. AFRS/RSOA also assists in implementation of new selection and classification assessment measures and processes. These activities are consistent with an operational mandate to support improving selection and classification systems (tests and processes) to optimize recruiting efficiency for Air Force Officer and Enlisted accessions while continuously adapting to changing population characteristics, training dynamics/criteria, and needs of the Air Force.

The AFPC/DSYX Contribution to Human Capital Management and Strategic Human Resources Management through Mission Alignment

DSYX makes contributions to the Air Staff by following the mission as tasked by AFMAN 36-2664:

- Provide technical guidance to and consult with AF/A1PT in identifying and overseeing strategic human resource capital initiatives.
- Support human capital studies and research to support decision-making involving recruiting, selection, classification, promotion, utilization, and retention.
- Coordinate changes to Air Force Officer and Enlisted Classification Directories (AFOCD & AFECD).
- Support revision and validation of the Air Force Officer Qualifying Test (AFOQT), the PCSM, and the Test of Basic Aviation Skills (TBAS).
- Conduct development, validation, and revision of tests and assessments.
- Evaluate enlistment and commissioning standards.
- Provide technical oversight of operational selection, classification, utilization, promotion, and proficiency testing and assessments to ensure they meet professional and legal standards.
- Technically review requests to develop/implement new tests/assessments.
- Manage the Applied Performance and Assessment Testing Center at Lackland AFB, TX.

DSYX makes contributions to the Air Force Personnel Center by following the mission as tasked by AFPC Mission Directive 37, 2003 [1-up]:

- Manage and operate Air Force military personnel data and information systems, execute policies that govern active duty accessions, testing, classification, assignments, personnel record systems, and personnel assessment.
- Manage and operate Air Force civilian personnel data and information systems and personnel assessment programs.

The DSYX Testing Toolbox

General Ability/Aptitude Tests

Air Force Officer Qualifying Test (AFOQT). The AFOQT is used to help select candidates for officer commissioning programs and to classify commissioned officers into utilization specialties

such as manned aircraft pilot, Remotely Piloted Aircraft (RPA) pilot combat system operators, air battle manager, or technical. Air Force Officer Qualifying Test scores are also used as a quality metric in the integrated officer classification model. The AFOQT is available in two versions (Form T1 and T2). Each version consists of 12 subtests. Subtests are used to compute one or more of the five aptitude composites. Scores on the subtests relate to performance in certain types of training. AFOQT composite scores are reported in percentiles.

Armed Services Vocational Aptitude Battery (ASVAB). The (ASVAB) evaluates specific aptitude areas and provides a percentile score related to requirements for selecting and classifying individuals for the Armed Services. There are two ASVAB testing programs—Student and Enlistment. The Student Testing Program applies to ASVAB testing in educational institutions such as high schools and vocational trade schools. The Enlistment Testing Program applies to Armed Services Vocational Battery testing in authorized accessions testing facilities such as Military Entrance Processing Stations (MEPS) and Military Entrance Test Sites (METS). The Army is the executive agent for the overall ASVAB Testing Program. The Defense Personnel Assessment Center in the Office of People Analytics is the executive agent for the ASVAB. The Air Force computes four training classification composite scores for the ASVAB: Mechanical (M), Administrative (A), General (G), and Electronics (E). These scores are predictive of training success in a variety of military occupations.

Electronic Data Processing Test (EDPT). The EDPT evaluates the basic ability to complete formal courses for programming electronic data processing equipment. The EDPT is a multiple-choice test that contains measures of verbal ability, symbolic reasoning, and arithmetic reasoning. It is used to screen and select Airmen for career fields requiring this ability. It is available by paper-and-pencil and electronically on the Personnel Testing Station[1] platform.

Vocational Interests

Air Force Work Interest Navigator (AF-WIN). The AF-WIN is an internet-delivered interest inventory that matches examinees' interests on the dimensions of functional communities, job contexts, and work activities to AFSC job profile markers to identify their "best fit" Air Force Specialties. It takes 15-20 minutes to complete with the examinee indicating level of interest on a 5-point scale for 52 items. There is a version of the AF-WIN for enlisted AFSCs and two officer versions. One officer version is designed for use at the beginning of college to help examinees plan their curriculum to include coursework required for particular AFSCs. The second version is for use closer to commissioning when finalizing the AFSC assigned to a cadet upon commissioning.

Personality

Tailored Adaptive Personality Assessment System (TAPAS). The TAPAS uses a trait taxonomy that assesses facets of the Big Five personality factors using a multidimensional

[1] The Personnel Testing Station was formerly called the Test of Basic Aviation Skills test station.

pairwise preference (MDPP) format. The assessment requires about 30 minutes to complete. It is completed by all new recruits at the Military Entrance Processing Station at the same time they complete the Armed Services Vocational Aptitude Battery. It is also administered on the Personnel Testing Station platform for selected retraining AFSCs.

Self-Description Inventory (SDI). The SDI was first implemented on AFOQT Form S as a 220 item, trait-based personality assessment of the Big Five personality domains and two Air Force related scales (Team Orientation and Service Orientation). Factor analyses of SDI item content revealed broad six domains encompassing the Big Five domains plus Machiavellianism, with subsequent factor analyses of domain content revealing a total of 20 narrower trait facets. The AFOQT Form T version of the SDI contains 240 items that assess the Big Five personality domains and Machiavellianism and 30 underlying facets.

Although the SDI was initially developed for the U.S. Air Force, a collaborative initiative with allied forces led to adaptations of the SDI for research purposes in the militaries of Canada, United Kingdom, New Zealand, and Australia.

Miscellaneous/Specialty

Test of Basic Aviation Skills (TBAS). The TBAS is a battery of scognitive, multi-tasking, and psychomotor ubtests administered on a computer test station. Examinees are required to respond to computerized tasks using a keypad, joysticks, and foot pedals. The TBAS includes subtests measuring psychomotor coordination, cognitive abilities, and multi-tasking capabilities. A pilot candidate's AFOQT Pilot composite score (or, where applicable, Enlisted Pilot Qualifying Test (EPQT) score) and Federal Aviation Administration certified flying hours are combined with the TBAS measurements to formulate a PCSM score. Manned aircraft Pilot and RPA pilot selection boards receive each candidate's PCSM composite score on a percentile scale of 1 to 99. PCSM assists pilot selection boards to select candidates most likely to successfully complete undergraduate pilot training.

Air Traffic Scenarios Test (ATST). The ATST is part of the classification screening process for candidates for the enlisted Air Traffic Control (ATC) AFSC. The Air Traffic Scenarios Test consists of simulated Air Traffic Control scenarios where the examinee is scored on how effectively they manage the departure, landing, tracking, etc. of aircraft with minimal safety violations. The test is administered on the TBAS testing platform and takes about an hour to complete.

Multi-Tasking Test (MTT). The MTT measures the ability to shift attention from one task to another over a short period of time. The test includes four component tasks: Math, Visual, Memory, and Listening. In the math task, participants add three-digit numbers. In the memorization task, a list of letters is initially presented and then disappears; after a delay, a probe letter is presented and participants indicate whether or not the probe letter was included in the list. In the listening task, participants respond with a mouse click when they hear a high-pitched tone and ignore a low-pitched tone. Finally, in the visual monitoring task, a needle moves from right to left across a display resembling a fuel gauge and the goal is to reset the needle when it nears the end of the display. The test is administered on the TBAS testing platform and takes about 45 minutes to complete.

The DSYX Expertise and Resources Toolbox

Staff Expertise

- Test Development/Validation – Professionals in the AFPC/DSYX team have decades of experience in item writing, item selection, scale development, test development, and test validation. Current AFPC/DSYX team members have experience developing DoD tests such as AFOQT, ASVAB, SDI, and AF-WIN. In addition, the team has experience in commercial test development including globally-recognized tests such as the Wechsler scales, the Beck inventories, and employee selection tests such as the Watson-Glaser Critical Thinking Appraisal and the Bennett Mechanical Comprehension Test.
- Predictive Model Development/Validation – Numerous occupational-specific predictive models have been developed by AFPC/DSYX using pre- and post-accession tests. Numerous empirical and regression-based formulas to predict important performance-based outcomes have now been operationalized for selection and classification purposes.
- Job/Occupational Analysis – AFPC/DSYX members have extensive expertise in job/occupational analysis to include task, trait, and competency analysis. The results of numerous AFPC/DSYX-based job analysis studies are now used in developing predictive models, responding to career field inquiries, and setting standards for classification (e.g., based on ASVAB profiles).
- Vocational Interest – AFPC/DSYX personnel have dnlisted- and officer-level vocational interest inventories. The tools developed by AFPC/DSYX have moved beyond traditional, generic vocational interest inventories and are specific to Air Force occupational specialties. The inventories provide information on the ideal match between a potential recruit and an occupational specialty and provide guidance to the examinee regarding the cognitive and physical requirement for the job.
- Job Satisfaction – DSYX personnel hace conducted studies of job satisfaction using USAF Occupational Analysis (OA) data and internally-developed surveys to determine if DSYX tests and/or predictive models are contributing to improved satisfaction.
- Structured Interviews – AFPC/DSYX has worked with USAF career fields to create structured interviews, structured interview guides, and video-based instructions for conducting valid structured interviews.
- Ethics/Integrity – AFPC/DSYX staff members have extensive experience in ethical behavior, integrity, and counterproductive behavior. AFPC/DSYX has developed integrity tests and valid tests designed to detect the propensity to engage in counterproductive behavior.
- Realistic Job Preview Creation – AFPC/DSYX staff members have extensive expertise in developing realistic job preview videos based on subject matter expert (SME) input video-based interviews.
- Leadership – AFPC/DSYX staff members have extensive expertise in assessing theories/models of leadership competencies and in the evaluation of leadership potential to help senior leaders attract, develop, and retain talent to effectively and efficiently accomplish mission requirements. The expertise encompasses experiences gained through work in academia, private industry, and military/government, which aid in providing customers with valuable tools, analysis, and innovative insights designed to improve organizational performance.

Contractor Expertise

Consulting Firms. AFPC/DSYX has had the opportunity work with the most well-known consulting firms in industrial and organization psychology and government research. In addition, DSYX has been able to contract out some work to the most recognized experts in their respective fields, including former presidents of the Society of Industrial and Organization Psychology (SIOP) and leading authors in academia and cutting-edge commercial innovation.

Forward Looking: The Future of AFPC/DSYX

Increased Effort to have AFPC/DSYX Expertise, Services, and Interventions Recognized throughout the Air Force

Recent efforts by AFPC/DSYX have improved the visibility of the branch throughout the Air Force. Specifically, efforts to educate Career Field Managers (CFMs) on the tools and services provided by DSYX have resulted in operational Predictive Success Models for numerous career fields and expansion of the use of existing tests for selection and classification purposes. In addition, updated internal marketing materials (e.g., slide decks, tri-fold brochures) are being prepared to provide additional exposure for the beneficial offerings of AFPC/DSYX. Finally, high-profile attention to quality products such as AF-WIN are providing additional recognition for how DSYX can provide high-quality and cost-effective services to the Air Force. Additional efforts will need to be expended in this area in order for AFPC/DSYX to continue to thrive as a valuable internal asset.

Improved Technology

Recent and future advances in available technology will provide AFPC/DSYX with the capability to provide services and tools in a more efficient manner. Examples include item-banking, a combined test-development and test-delivery platform, and even sophisticated tools such as text analysis.

Improved Access to Data

Current processes to procure and process necessary data (e.g., test scores, training grades) are somewhat inefficient and hinder the efficiency and effectiveness of the branch. Future enhancements are being vetted and implemented to automate and streamline the process. This will allow DSYX to provide real-time decision support to internal clients and will improve the speed in which AFPC/DSYX can build the tests and tools required for effective selection and classification purposes.

Exiting the Operational Testing Domain

DSYX historically has been involved in many aspects of operational testing (e.g., test delivery, scoring, coding) which limits the time and resources available to devote to true mission-specific activities. Current efforts are being conducted to ensure a more efficient hand-off from AFPC/DSYX to the operational entities after successful development of tests and selection/classification tools.

Repeatable and Scalable Processes

AFPC/DSYX is currently striving to develop repeatable (e.g., consistent analyses, similar technical report templates) and scalable analyses and processes (e.g., processes that can be applied to large and small requests throughout the Air Force). This Guide is one small step in achieving that goal.

1.0 ETHICAL AND LEGAL CONSIDERATIONS IN PERSONNEL TESTING

1.1 Introduction

The purpose of this chapter is to discuss what the author believes are the best practices for understanding and evaluating critical legal issues relating to adverse impact. The chapter is divided into six sections including:

- Section 1: A comparison of adverse impact to other forms of discrimination in facial discrimination, disparate treatment, and the pattern or practice of discrimination.
- Section 2: An overview of major court rulings related to adverse impact.
- Section 3: A discussion of major issues relating to the *Uniform Guidelines on Personnel Selection* (or simply *Guidelines* (1978), including those Guidelines that are still relevant as compared to those that have been struck down in the courts.
- Section 4: An examination of how to create and validate test batteries.
- Section 5: How and when to employ content validity strategies.
- Section 6: A review of court rulings related to minimizing adverse impact.

At the outset, it is important to recognize that the Federal antidiscrimination laws do not apply directly to military selection issues. That said they are important insofar as the regulations that govern some of these laws, most notably adverse impact, play an important role in determining if a selection test or other selection procedures are valid. Valid tests are needed in order to successfully predict positive outcomes in personnel selection.

1.2 Section 1: Overview of Types of Discrimination

The purpose of this section is to understand what adverse impact is in the context of two other forms of discrimination; facial discrimination and disparate treatment. Additionally, adverse impact will be compared to the *pattern or practice* of discrimination, which is a form of disparate treatment that has been frequently confused with adverse impact because both require statistical proofs of discrimination. However, as we will witness, the types of statistics used for adverse impact and pattern or practice are different from each other.

1.2.1. Facial Discrimination

In an earlier era (e.g. the 1800s), it was common to see signs such as "no blacks, Jews or women allowed." Signs such as these illustrate what facial discrimination is; exclusion on its face of individuals because of their race, color, sex, religion or national origin. The motive for facial discrimination in the earlier era was simple; certain groups or classes of individuals did not want to associate with other groups or classes of individuals.

The Civil Rights Act of 1964 (CRA-64) made it illegal to facially discriminate based on race or color for **any reason** but permitted facial discrimination based on sex, age, religion or national origin if it could stand the Bona Fide Occupational Qualification (BFOQ) defense, which requires proof that it is reasonably necessary to exclude all or most members in a given class.

To illustrate, in *Dothard v. Rawlinson*, a case to be discussed again in Section 2 below, women were facially excluded from being guards in an all-male maximum security prison in which 20 percent (%) of the inmates were sex offenders. The Supreme Court favored the prison on grounds that having women in the presence of male sex offenders served as a threat to prison safety.

In general, the BFOQ defense has succeeded when the issue is workplace or customer **safety** and customer **privacy**, but **never** for customer preference. To illustrate, airlines may exclude pregnant flight attendants at some point in the pregnancy due to implied threat to passenger safety (e.g., *Levin v. Delta Airlines*, 1984), all-male bathhouses may exclude female janitors for privacy reasons (*Brooks v. AFC Industries*, 1982), and all-female reformatories may exclude male counselors because males may be perceived as threatening by female patients (*Torres v. Wisconsin*, 1988). However, airlines may not exclude male flight attendants based on passenger preference (e.g., *Diaz v. Pan American*, 1971) or because of a need to preserve a "sexy image" (e.g., *Wilson v. Southwest*, 1981), as neither motive is related to the main job of airlines, which is to transport passengers.

1.2.2. Disparate Treatment

The motive to discriminate is obvious in facial discrimination (i.e., true on its face). When disparate treatment occurs, it is usually the case that a false reason is used as a pretext for refusal to hire or promote an individual. For example, the author, while consulting for a well-known high-tech company in the 1980s, faced a situation in which three black employees took sick leave to attend the World Series in Atlanta, Georgia. Their manager subsequently terminated them on grounds that they violated the sick leave policy. However, there was direct evidence of racial discrimination by the manager (witnesses) and indirect evidence of racial discrimination (nobody had ever been terminated for violating the policy). As a result, the manager was terminated and the three employees received remuneration.

In an earlier era, plaintiffs were required to show direct evidence of discrimination (e.g., witnesses, documents, etc.) in order to prevail in a disparate treatment claim. Plaintiffs rarely had such smoking gun evidence. Then, the Supreme Court altered the rules for proving disparate treatment in *McDonnel Douglas v. Green* (1973).

Percy Green engaged in civil rights activism, which is legal behavior. During a layoff, Green engaged in two illegal acts against the company; a stall-in in which cars were prevented from entering the plant and a lock-out in which cars were prevented from leaving the plant. The company then advertised for his prior job and Green applied but was not rehired. The lower courts ruled against Green because he had no direct evidence of racial discrimination. The Supreme Court then outlined a novel three-phase procedure for proving disparate treatment. Green lost anyway, but the more important precedent was establishing the new scenario. Table 1 illustrates the three-phase scenario for proving disparate treatment in the *McDonnel Douglas* case.

Table 1. The McDonnel Douglas Scenario

Phase 1 (prima facie phase)	Plaintiff is a protected class member; he applied and was qualified for the job; he was passed over; and the search continued
Phase 2 (defense phase)	Defendant articulates (explains) without having to prove a nondiscriminatory reason for not re-hiring
Phase 3 (pretext phase)	Plaintiff must prove with direct or indirect evidence that articulation provided is a pretext for discrimination

In Phase 1, Green presented prima facie evidence that he was a protected class member (he was black), that he applied and was qualified for the job (which he had performed before), that he was passed over (which he was), and that the search continued (which it did). In Phase 2, the company articulated without having to prove that Green was not rehired because of his participation in illegal activities during the layoff. In Phase 3, Green failed because the company had no history of racial discrimination and it did not rehire anyone, white or black, who had participated in the illegal acts during the layoff.

1.2.3. Adverse Impact

Adverse impact is illustrated by the aftermath of the 15th Amendment, which granted former black slaves the right to vote. Many municipalities created arbitrary requirements to exercise that right such as poll taxes and land ownership. Such requirements limited blacks from voting at a much higher rate than whites. Ultimately, the 26th Amendment eliminated requirements to vote (other than citizenship) in 1971. As we will witness below, an analogous set of events occurred in *Griggs v. Duke Power* (1971) after CRA-64 was enacted. More generally, selection criteria involving cognitive tests tend to adversely impact blacks and height and weight requirements tend to adversely impact women. Proof of adverse impact requires use of either applicant flow data, (actual selection rates) or demographic distributions that imply what these selection rates would be.

In *Griggs*, the causes of adverse impact were two cognitive tests and the requirement to possess a high school diploma. The tests excluded a much higher percentage of **actual** black applicants than white applicants and the diploma was **presumed** to do likewise because demographic statistics revealed that the graduation rate in North Carolina at the time was nearly three times higher for whites than blacks.

The adverse impact scenario is illustrated in Table 2. Phase 1 requires proof that a test or other selection procedures exclude an appreciably higher percentage of one protected group as compared to another. If proven, the defendant must prove that the test or other selection procedures are job related and consistent with business necessity in Phase 2. If successful, the plaintiff(s) must prove there are other tests or selection procedures that are equally as valid and produce less or no adverse impact.

Table 2. The Adverse Impact Scenario

Phase 1 (prima facie phase)	Plaintiff(s) must prove that a test or other selection procedures exclude a significant higher percentage of one protected group compared to another protect group
Phase 2 (defense phase)	Defendant must prove that their test or other selection procedure is job related and consistent with business necessity
Phase 3 (pretext phase)	Plaintiff must prove there are alternative tests or other procedure that are as valid but produce less or no adverse impact

1.2.4. Pattern or Practice

Proof of adverse impact involves **applicant flow** statistics (i.e., selection rate disparities between actual groups of applicants or presumed selection rate disparities based on demographic statistics), whereas pattern or practice discrimination involves **stock** statistics. (Stock statistics are simply descriptive statistics that indicate the current state of a situation, such as percent of minority vs. nonminority members in an occupation at a given moment in time.) There are two types of stock statistics, namely cross-job disparities, which reflect differences across jobs and are illustrated in *International Teamsters v. US* (1977) and composition disparities, which reflect differences rates of groups in a job vs. rates at which they are available in the workforce, as illustrated in *Hazelwood School District v. US* (1977).

Cross-job disparities: *Teamsters* featured two bus driving jobs. One of them involved over the road routes (first job type) and was higher paying than the other one, which involved shorter local routes (second job type). Blacks and Hispanics were permitted to drive the local routes but were facially excluded from the longer routes. The Teamsters had no nondiscriminatory explanation for the cross-job disparity and they lost at the Supreme Court level. A similar scenario later occurred in *Wards Cove v. Atonio* (1989); however, in my opinion, the Supreme Court mistakenly treated it as an adverse impact case and not a pattern or practice case.

Composition disparities: In *Hazelwood*, the issue was composition disparities between the workforce and those in the labor pool who were qualified and available. There was a lower percentage of black teachers than white teachers when the comparison included the entire county as well as the local Hazelwood district. However, the county disparity was statistically significant, whereas the local disparity was not. The Supreme Court favored the local comparison and the government lost. Additionally, Hazelwood had an additional consideration in that teachers did not generally want to travel the longer distance from St. Louis to the Hazelwood district. A more relevant example for the Air Force would be if 13 percent of women receive civilian pilot licenses nationally, the percentage selected to be pilots in the Air Force is expected to be 13 percent, as well. To the extent that percentage deviates in the Air Force, this could constitute a composition disparity. Of course, whether this ration is statistically significant would depend on the overall numbers of women in the Air Force.

To summarize, seven major points were made in this section:

1. Facial discrimination is prohibited based on race or color but is permitted if defensible via BFOQ based on sex, religion or national origin.
2. The BFOQ defense is most likely to succeed if there are reasonable concerns related to safety and/or privacy, but never for preference.
3. The *McDonnell Douglas* scenario forces defendants to explain why a given selection decision is made, thus setting up the possibility for plaintiff to prove pretext.
4. Since there is rarely a smoking gun (direct evidence) of discrimination, most plaintiffs use indirect evidence.
5. Adverse impact involves flow statistics, whereas pattern or practice involves stock statistics.
6. The adverse impact proof may involve selection rate disparities among actual applicants and/or presumed selection rate disparities based on demographic statistics.
7. The pattern or practice statistics involve cross-job disparities or composition disparities; both are stock statistics and are decidedly different than flow statistics.

1.3 Section 2: Landmark Adverse Impact Rulings

Table 3 depicts seven Supreme Court landmark rulings on adverse impact. The rulings in *Griggs* through *Connecticut v. Teal (*1982) established early precedents that were temporarily altered in *Watson v. Ft. Worth Bank* (1988) and *Wards Cove* before being reestablished for all intents and purposes in the Civil Rights Act of 1991 (CRA-91). More specifically, in *Watson*, only eight justices heard the case and all eight agreed on one issue but were split 4-4 on a second issue related to the Phase 2 defense to adverse impact. Then, a year later, Justice Kennedy joined the Court, forging a majority of five that temporarily altered the Phase 2 defense to adverse impact established in pre-*Watson* cases before Congress reestablished the pre-*Watson* defense in CRA-91.

Table 3. Supreme Court Landmark Rulings

Griggs v. Duke Power (1971)	Successful challenge to arbitrary use of high school diploma and cognitive ability tests
Albemarle v. Moody (1975)	Faulty validation study for cognitive ability tests is struck down; pretext phase is added
Washington v. Davis (1976)	Civil Service tests for hiring police officers is upheld because they predict training school performance
Dothard v. Rawlinson (1977)	Adverse impact on women based on height and weight as surrogate for strength is struck down
Connecticut v. Teal (1982)	Any hurdle of a multiple hurdle selection procedure must be validated if there is adverse impact regardless of what the bottom-line numbers are
Watson v. Ft. Worth Bank (1988)	Subjective causes of adverse impact affirmed; however, a plurality of justices favor altering the Phase 2 defense to adverse impact.
Wards Cove v. Atonio (1989)	Prior precedents for defending against adverse impact are temporarily reversed and then recovered in CRA-91.

1.3.1. Griggs v. Duke Power (1971)

As noted in Section 1, the facts in *Griggs* were analogous to the events following enactment of the 15th Amendment, which gave blacks the right to vote. Duke Power historically hired blacks for lower-wage labor jobs but blacks were facially excluded from higher-wage operations jobs. In 1955, the company adopted a High School (HS) diploma requirement for entry into operations jobs. Then, on July 2, 1965, the **very day** Title VII became law, all new operations employees were required to possess the diploma and pass two cognitive tests (the Bennett Mechanical Aptitude Test and Wonderlic Personnel Test). Whites without diplomas hired before 1955 could transfer to operations by passing the cognitive tests. Interestingly, before 1955, whites without diplomas were routinely promoted to operations. Thus, there was evidence before July 2, 1965 that both the diploma and cognitive abilities were **arbitrary** requirements. Thus, Duke Power gave the appearance of intentionally limiting the number of blacks in higher-level operations jobs. Despite appearances, the lower courts found no motive to discriminate.

The Supreme Court ignored the disparate treatment claim and focused only on the **effects** of the challenged practices. The cognitive tests excluded 94% of **actual** black applicants but only 42% of **actual** white applicants. The diploma had a similar but **presumptive** effect as demographic data revealed that the graduation rate in North Carolina at the time of the lawsuit was 34% for whites but only 12% for blacks.

Duke Power lost on both counts. However, at the end of the day, this was mainly a cognitive ability testing case. For its part, Duke Power relied on CRA-64 statutory language making it legal to use, "professionally developed tests" such as the Bennett and Wonderlic tests. However,

the 1966 Equal Employment Opportunity Commission (EEOC) Guidelines defined a "professionally developed test" as one that:

> [F]airly measures the knowledge or skills required by the particular job or class of jobs which the applicant seeks, or which fairly affords the employer a chance to measure the applicant's ability to perform a particular job or class of jobs.

The lower courts ignored this guidance but Justice Burger, speaking for a unanimous Supreme Court, ruled that Title VII covers the "**consequences** of employment practices, not simply the **motivation**" of employers. He then wrote two critical words that are etched in adverse impact case law; that if the plaintiff proves adverse impact, the defendant must prove there is a "**manifest relationship"** between the challenged practice and the employment in question.

1.3.2. Albemarle v. Moody (1975)

Albemarle was *Griggs* revisited but with a major twist. The company used the HS diploma and the Bennett and Revised Beta Examination. Aware of the *Griggs* ruling, the company hired an expert to correlate test scores with job performance ratings (i.e., a criterion validity study). However, it was a hasty effort conducted shortly before trial. Indeed, the expert was not even on site and the validation study was packed with defects. Citing the revised 1970 EEOC *Guidelines*, the Supreme Court defined how a **manifest relationship** should be proven. Accordingly:

> The message of these Guidelines is the same as that of the Griggs case—that discriminatory tests are impermissible unless shown by professionally acceptable methods to be predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are evaluated.

Interestingly, Justice Burger, who coined the phrase "manifest relationship" in *Griggs*, was the only justice who dissented in *Albemarle*. The *Albemarle* Court then outlined the Phase 3 requirement on alternative practices with less or no adverse impact. Subsequently, the *Uniform Guidelines* were adopted in 1978 based primarily on the *Griggs* and *Albemarle* rulings.

1.3.3. Washington v. Davis (1976)

Washington was a 5th amendment challenge to Civil Service Test 21 used for hiring purposes by the Washington, D.C. Police Department. The Supreme Court ruled that adverse impact cannot be challenged via constitutional claims, but nevertheless, the Court evaluated Test 21, a measure of verbal skills, and concluded it was valid because it predicted training school performance. Thus, in professions such as law enforcement, where formal academy training is a precursor to a formal hiring decision, the *Washington* ruling implies it is permissible to statistically correlate test scores with training performance to establish **criterion validity**. This same principal would apply to pilot performance, or any situation where training is necessary to the job performance in question. Note that the interpretation here is that the validation is based on job related training, not simply training in and of itself. For example, simple completion of training vs. training attrition would not be considered an effective criterion measure in and of itself.

1.3.4. Dothard v. Rawlinson (1977)

Prior to its victory on the BFOQ defense for facial discrimination of women, the State of Alabama lost an adverse impact challenge to a minimum height/weight requirement (regardless of sex). Demographic data clearly indicated a presumptive adverse impact effect on women. The State argued that height/weight criteria were indicative of (or a surrogate for) strength, which is a requirement for prison guards. The Supreme Court ruled that if strength was job related, the State needed to create and validate a direct measure of strength. The moral of the *Dothard* case is that it was easier for the State of Alabama to facially exclude all women than it was to eliminate all people who did not meet the minimum height/weight requirement.

1.3.5. Connecticut v. Teal (1982)

In *Connecticut*, provisional promotions for several black candidates were rescinded after they failed a written test that was the first step in a multiple hurdle process. Interestingly, after all hurdles were completed, a higher percentage of blacks (22.9%) than whites (13.5%) were promoted. However, the passing rate for blacks on the written test was only 68% relative to whites. Connecticut appealed to Section 1607.4(C) of the *Uniform Guidelines*, which has two provisions that state:

> [1] If … the total selection process for a job has an adverse impact, the individual components of the selection process should be evaluated for adverse impact.
>
> [2] If … the total selection process does not have an adverse impact, the Federal enforcement agencies … will not expect a user to evaluate the individual components for adverse impact, or to validate such individual components, and will not take enforcement action based upon adverse impact in any component of that process.

In other words, that adverse impact for the total selection process **implicates** all steps in a multiple hurdle but absence of such bottom-line adverse impact **insulates** all hurdles. A unanimous Supreme Court affirmed the implication provision but a 5-4 majority struck down the insulation provision. In short, the 5-4 majority concluded that **all** requirements in a multiple hurdle system must be free from adverse impact on protected group members, or job-related, regardless of the bottom-line results.

The *Connecticut* ruling will be discussed further in Section 3 below. For present purposes, it is critical to recognize that most selection procedures have multiple steps or hurdles and that regardless of what the end result is, each step should be evaluated for adverse impact and the possible need for validation. Thus, if the Air Force uses a series of pilot selection tests, applicant flow disparities should be assessed for each test as well as for the combined, bottom line results.

1.3.6. Watson v. Fort Worth Bank (1988)

Watson and *Wards Cove* were strange bedfellows. The central issue in *Watson* was more closely related to traditional *Griggs–Albemarle* principles, whereas the central issue in *Wards Cove* was more closely related to *Teamsters–Hazelwood* principles. Together, they formed a two-act play, and when the curtain fell, pre-*Watson* traditions were dramatically, albeit temporarily, altered.

In the *Watson* case, Clara Watson, a black woman, was passed over for promotion four times, each time in favor of a white applicant and each time based on **subjective** ratings by white supervisors. Watson challenged subjective ratings of: (1) job performance, (2) interview performance, and (3) past experience. It was unclear how these ratings were combined, but there was clearly bottom-line adverse impact

1.3.7. Wards Cove v. Atonio (1989)

In *Wards Cove*, two Pacific Northwest salmon packing companies had a hiring hall arrangement for unskilled salmon packers, but used different procedures to hire skilled workers (e.g., machinists). Skilled workers were paid, fed, and housed better than the unskilled packers. Eskimos and Filipinos were overrepresented in the unskilled jobs and underrepresented in the skilled jobs. Thus, on its face, *Wards Cove* was *Teamsters* revisited. As in *Teamsters*, minorities were congregated in a less desirable job and whites in more desirable jobs.

The plaintiffs claimed both pattern or practice and adverse impact. Eskimo and Filipino salmon packers had previously made both claims based on cross-job disparities in *Domingo v. New England Fish* (1984) and the 9th Circuit virtually ignored the adverse impact claim and ruled there was a pattern or practice violation. However, given a second opportunity, the 9th Circuit added language that subjective hiring procedures (i.e., separate hiring procedures for skilled vs. unskilled jobs, nepotism, word of mouth recruitment, and "vague, subjective hiring criteria") were involved. Thus, they upheld the district court's rejection of pattern or practice, but reversed the district court's rejection of adverse impact, ruling that there can be subjective causes of adverse impact.

For its part, the Supreme Court seized on the subjectivity ruling and used it to reiterate Justice O'Connor's three provisions in the *Watson* case. CRA-91 (see Table 4) accepted the first two provisions (identification and causation) but overturned on the third provision (burden of production) such that if adverse impact is proven, the burden shifts to the defendant to prove that the cause(s) of adverse impact is/are job related and consistent with business necessity. The meaning of "consistency with business necessity" is an issue that has yet to be clarified. Generally, though, courts have accepted job relatedness as sufficient to satisfy the Phase 2

Table 4. Justice O'Connor's Proposed Provisions

Identification Provision	Plaintiff must **identify** specific cause(s) of adverse impact unless the selection process cannot be disaggregated
Causation Provision	Statistical disparities must be sufficiently large enough to establish **causation** of adverse impact
Burden of Production	Defendant must **articulate** without having to prove a nondiscriminatory reason for using the selection procedure(s) (as in *McDonnel Douglas v. Green*, 1973)

The first two provisions were not revolutionary. In fact, all the cases from Griggs to Connecticut satisfied both provisions. Provision 3, on the other hand, would effectively make the defense to adverse impact synonymous with the defense in disparate treatment as in McDonnel Douglas.

requirement. That is, the policy or practice in question must be directly associated with the skills required to needed to perform the job. This is opposed to general business or organizational outcomes such as reduction of training costs or minimizing attrition. Phase 3 of the adverse impact scenario (alternatives with less or no adverse impact) remained untouched.

To summarize, 11 major points were made in this section:

1. The *Griggs* ruling ignored what was likely a motive to reduce minority participation by focusing on the effects of selection procedures rather than the motivation of employer.
2. The *Griggs* ruling mandated that tests or other selection procedures that cause adverse impact must be job related.
3. The *Albemarle* ruling made it clear that poor, hastily conducted validity studies will not be sufficient to prove job relatedness.
4. The *Albemarle* ruling also created the pretext phase in adverse impact allowing plaintiffs to prove there are alternative tests or selection procedures that are as or less valid that produce less or no adverse impact.
5. The *Davis* ruling made it plausible to use training scores in criterion validity study in those professions that require training prior to beginning working.
6. The *Dothard* ruling showed that it was easier for a prison system to facially exclude all women via a BFOQ defense than all people that did not meet a height/weight criterion in an adverse impact defense.
7. The *Connecticut* ruling made it necessary to examine all components of a multiple hurdle selection procedure for adverse impact regardless of the bottom-line results.
8. The *Watson* ruling made it legal to challenge subjective selection criteria with an adverse impact challenge.
9. The *Watson* ruling also led to a plurality opinion that would alter the defense to adverse impact to mimic disparate treatment claims.
10. The *Wards Cove* ruling temporarily replaced the *Griggs-Albermarle* defense with the *McDonnell Douglas* defense.
11. The *Griggs-Albemarle* defense was, for all intents and purposes reestablished in the Civil Rights Act of 1991.

1.4 Section 3: Uniform Guidelines on Personnel Selection Procedures

As noted in Section 2 above, the *Griggs* ruling followed the prescriptions of the 1966 EEOC Guidelines and the *Albemarle* ruling followed the prescriptions of the 1972 EEOC Guidelines. Indeed, by the early 1970s, five agencies had guidelines related to selection procedures including the Civil Service Commission (CSC), Civil Rights Commission, Department of Justice (DOJ), Department of Labor (DOL) and, of course the EEOC. The *Uniform Guidelines on Employment Selection Procedures* (or simply the *Guidelines*) were written to unify these interests. They were formally adopted in 1978. Primary ownership fell to the EEOC as a result of President Carter's Reorganization Plan #4 (1978) but other agencies maintained an active interest, most notably the DOJ and the Office of Contract Compliance Programs (OFCCP) of the DOL, which enforces Executive Order 11246 on affirmative action.

In 1979, the EEOC and other interested agencies adopted 90 questions and answers to "clarify and provide a common interpretation" of the *Guidelines*. In 2004, the EEOC and OFFCP

updated Question 15 of the 1979 definition of **applicant** to address emerging issues associated with Internet recruitment. However, there have been no changes to the *Guidelines* themselves since their adoption in 1978. Consequently, employers face the challenge of interpreting outdated regulations that courts rely on to evaluate adverse impact claims. Exacerbating this challenge is the fact that some *Guidelines* have been struck down in court rulings as, for example, the insulation provision in the *Connecticut* case. The purpose of this section is to provide a blueprint for navigating the *Guidelines* in the face of these challenges. The most important considerations are understanding the contribution of the psychological profession to the *Guidelines* and understanding how some of the critical guidelines have fared in court.

1.4.1. Role of the Psychological Profession

The following is written into Section 1607.1(c) of the *Guidelines* on "relation to prior guidelines."

> These guidelines are based upon and supersede previously issued guidelines on employee selection procedures. These guidelines have been built upon court decisions, then previously issued guidelines of the agencies, and the practical experience of the agencies, as well as the standards of the **psychological profession**. These guidelines are intended to be consistent with existing law.

The authors of the *Guidelines* also recognized that new validation strategies would be developed over time. Thus, Section 1607.5(A) of the *Guidelines* pays respect to **new strategies** accepted by the **psychological profession** as follows in a section entitled "acceptable types of validity studies."

> For the purposes of satisfying these guidelines, users may rely upon criterion-related validity studies, content validity studies or construct validity studies, in accordance with the standards set forth in the technical standards of these guidelines, section 14 below. **New strategies for showing the validity of selection procedures will be evaluated as they become accepted by the psychological profession**.

The two most important authorities for keeping abreast of new strategies in validation research are the SIOP *Principals for the Validation and use of Employee Selection Procedures* (or simply the *Principals*) (5th Edition, 2018) and the *Standards for Educational and Psychological Testing* (or simply the *Standards*) authored by the American Psychological Association (APA), American Educational Research Association (AERA), and the National Counsel on Measurement in Education (NCME) (7th Edition, 2014). Both are freely available on the Internet.

From the viewpoint of the psychological profession, the *Guidelines* were outdated from the outset. Landy (1986) was among the first in the psychological profession to criticize the trinitarian view of validity in the *Guidelines* in which content, criterion, and construct validity are defined as having different purposes and criterion validity is viewed as the gold standard for proving job relatedness (see also Gutman, 2005; Landy, Gutman, & Outtz, 2010). Landy was particularly critical of Section 1607.C(1) of the *Guidelines* on "appropriateness of content validity studies." This section contains the following frequently cited warning against relying "solely or primarily" on content validity to support mental processes. Accordingly:

> A selection procedure based on inferences about mental processes cannot be supported solely or primarily on the basis of content validity. Thus, a content strategy is not appropriate for demonstrating the validity of selection procedures which purport to measure traits or constructs such as intelligence, aptitude, personality, common sense, judgment, leadership and spatial ability.

For reasons to be discussed shortly, this guidance is outdated and has been struck down by the courts. For present purposes, the important point to note is that blind adherence to the letter of the *Guidelines* could lead to unfortunate errors in decision making. For example, if a test is properly constructed based on job analysis data, it should stand up to a content validity study. But if it does not stand up to a criterion validity study, which can happen simply for technical reasons, it would be a mistake to discard the test when, in fact, by satisfying content validity principles, the test itself would, indeed be considered job related and consistent with business necessity.

1.4.2. Content Validity

The aforementioned guidance related to content validity and assessment of mental abilities is one of two critical *Guidelines* struck down by the courts, the other being the insulation provision in relation to the *Connecticut* case. Also important is the "80 percent" rule for proving adverse impact. The immediate discussion focuses on the content validity issue.

Despite the ominous warning in Section 1607.C(1), courts immediately and universally supported content validity as a viable method for proving job relatedness. The landmark ruling on content validity is *Guardians v. Civil Service* (1980). In *Guardians*, the defendants attempted to support strict rank ordering based on a content validity study. They lost but the 2nd Circuit Court outlined the following five steps for a content validity study to support strict rank ordering:

1. Suitable job analysis
2. Reasonable competence in test construction
3. Test content related to job content
4. Test content representative of job content
5. Scoring systems selecting applicants who are better job performers

These steps will be discussed in greater detail in Section 5 below. For present purposes, the defendants in *Guardians* failed Steps 2 and 3. Nevertheless, based on *Guardians*, other defendants successfully supported job relatedness and strict rank ordering with properly designed content validity studies.

For example, in *Gillespie v. Wisconsin* (1985), the 7th Circuit Court rejected the notion in Section 1607.C(1) that the *Guidelines* "prefer criterion-related validity." Citing Question 56 of the 1979 questions/answers, the 7th circuit ruled that "because the Guidelines describe the conditions under which each validity strategy is inappropriate, there is no reason to state a general preference for any one validity strategy." Moreover, citing Question 62 of the 1979 questions answers verbatim, the 7th circuit ruled that:

> Use of a selection procedure on a ranking basis may be supported by content validity if there is evidence from job analysis or other empirical data that what is measured by the selection procedure is associated with differences in levels of job performance.

Additionally, citing both the 1974 *Standards* and Anastasi's (1982) text on *Psychological Testing*, the 7th Circuit Court ruled that:

> [P]sychometricians who have studied employment tests have determined that criterion-related validation is often impracticable because of the limited numbers of employees available for test development and several measurement errors … neither the Uniform Guidelines nor the psychological literature express a blanket preference for criterion-related validity.

Subsequently, in *Police Officers v. City of Columbus* (1990), the Sixth Circuit Court of Appeals credited the 1987 *Principles* and accepted a content validity study to support strict rank ordering. Accordingly, "the SIOP principles provide: If selection instruments measure a substantial and important part of the job reliably, and provide adequate discrimination in the score ranges involved, persons may be ranked on the basis of its results."

The *Guardians* ruling was subsequently supported by other courts (e.g., *Association of Mexican-American Educators v. California,* 2000; *Bew v. Chicago,* 2001; *Brunet v. City of Columbus,* 1993; *Williams v. Ford Motors,* 1999), & *Gulino v. New York State* (2006).

In short, courts have generally accepted the use of content validity as a method for supporting job relatedness. In fact, in the author's opinion, content validity has become the validation method of choice mainly because of the technical difficulties involved in performing criterion validity studies, an issue to be discussed in greater detail in Section 4 below.

1.4.3. Multiple Hurdle and Multiple Selection Procedures

Recall that in *Connecticut*, the Supreme Court struck down the **insulation** provision, which had made it safe to not assess the individual components of a multiple hurdle if there was no bottom-line adverse impact for the total selection process. Thus, it is incumbent on employers to examine each step of the process. This need was further endorsed in CRA-91, which stipulates that if the specific components of the selection process cannot be disaggregated, this responsibility to defend the entire process falls to the defendant.

More generally, employers should evaluate all components of a selection process. It is critical to examine (1) how each component is scored, (2) how scores are combined, and (3) whether applicants are excluded by any single component that serves as a minimal qualification for the job. As most selection processes involve multiple selection components, employers are advised to understand the strengths and potential weaknesses of any component.

For example, lost in the headlines relating to subjective selection procedures in the *Watson* case, it was not clear how Fort Worth Bank analyzed the subjective ratings of job performance, interview performance, and past experience. It is clear that each component had four raters but it is not clear if the ratings were made independently or reliably. Nor is it clear how these scores

were combined. In the author's opinion, any process is suspect under these circumstances, regardless of whether the components are subjectively or objectively scored.

Gilbert v. Little Rock (1983) provides another good example of what **not** to do with multiple components. In *Gilbert*, a police department combined an objective pass/fail test with subjective oral interview scores. The pass/fail rates for blacks and whites were almost equal, but adverse impact occurred after Step 2 (subjective oral criteria). There was also evidence of tampering. Each year, objective (written) and subjective (oral) criteria were combined to form a final ranking. Then, promotions were made in rank order until all available opportunities were exhausted. Each year, many blacks were ranked, but none high enough to be promoted. Basically, the oral evaluations were manipulated so that blacks with high written scores had poor oral scores and blacks with low written scores had high oral scores. Thus, the court ruled that the scores were manipulated so that blacks high on one measure were low on the other and vice versa.

In short, the most worrisome issue in multiple hurdle/selection procedures is not being able to document how scores were obtained or if reliability assessments were made if subjective rankings were made. The employer must be able to define the entire process and be in a position to disaggregate and defend every hurdle.

1.4.4. The Four-Fifths (Or 80% Rule)

The four-fifths (or 80%) rule is a simple rule of thumb for computing adverse impact. For example, if the selection rate for whites is 70% (e.g., 70 of 100), the selection rate for minorities or women should be .80 × 70% = 56% (or 56 of 100 or higher); or as written in Section 1607.4(D) of the *Guidelines*:

> A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact.

The danger here is **not** in misunderstanding the 80% rule, but rather, in overemphasizing it relative to **statistical** significance or **practical** significance.

In that regard, there is a second part to Section 1607.4(D) that is often overlooked. Accordingly:

> Smaller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both **statistical** and **practical** terms or where a user's actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group. Greater differences in selection rate may not constitute adverse impact where the differences are based on small numbers and are not statistically significant.

Comprehensive discussions of the 80% rule compared with statistical and practical significance, including statistics for small and large sample sizes, are provided by Morris and Lobsenz (2000) and Siskin and Trippi (2005). Unfortunately, there have been no court cases demonstrating

specific examples of employers' actions having discouraged applicants. For present purposes, the focus is on regulatory language and court cases illustrating the danger in relying **exclusively** on the 80% rule.

The two best sources of information on practical and statistical significance are written in the answers to Questions 20 and 21 of the questions/answers regarding the *Guidelines*. Question 21 is entitled "why is the 4/5ths rule called a rule of thumb?" Two examples are given. The first example illustrates statistical and practical significance in situations where the 80% rule is **not** violated.

> …assume that nationwide statistics show that use of an arrest record would disqualify 10% of all Hispanic persons but only 4% of all whites other than Hispanic (hereafter non-Hispanic), the selection rate for that selection procedure is 90% for Hispanics and 96% for non-Hispanics. Therefore, the 4/5 rule of thumb would not indicate the presence of adverse impact (90% is approximately 94% of 96%). But in this example, the information is based upon nationwide statistics, and the sample is large enough to yield statistically significant results, and the difference (Hispanics are 2 1/2 times as likely to be disqualified as non-Hispanics) is large enough to be practically significant.

The second example illustrates that the 80% rule when violated in small sample sizes may not yield an inference of adverse impact.

> On the other hand, a difference of more than 20% in rates of selection may not provide a basis for finding adverse impact if the number of persons selected is very small. For example, if the employer selected three males and one female from an applicant pool of 20 males and 10 females, the 4/5ths rule would indicate adverse impact (selection rate for women is 10%; for men 15%; 10/15 or 66 2/3% is less than 80%), yet the number of selections is too small to warrant a determination of adverse impact.

Lastly, the answer to question 21 illustrates the so-called flip-flop rule as it is applied to small sample sizes. The answer assumed the data in Table 5below, where the selection rates are 20% for whites and 15% for blacks and the 80% rule is violated (.8 x 20 = 16%). However, if just one of the 17 blacks is shifted to hired than the black selection rate now equals the white selection rate.

Table 5. The Flip Flop Rule

Applicants	Not Hired	Hired	Selection Rate Percent Hired
80 Whites	64	16	20
40 Blacks	17	3	15
White Selection Rate....			20
Black Selection Rate....			15

The best example in overreliance on the 80% rule is in *Bew v. Chicago* (2001), in which a test to certify probationary police officers administered to 5,191 applicants yielded pass rates of 99.96 and 98.24%, respectively for whites and blacks. Thus, the pass rate for blacks was 98.28% that of whites; well within the boundaries of the 80% rule. However, critically, 32 of the 33 applicants who failed were blacks, and adverse impact was inferred based on a test of independent proportions revealing a nearly five standard deviation difference between these failure rates.

Similarly, in *Isabel v. City of Memphis* (2005), the 80% rule was violated with its original cutoff score of 70. So the cutoff score was reduced to 66 yielding a pass rate of 46 of 63 for blacks (74.6%) and 51 of 57 for whites (89.5%). The city felt safe because the pass rate for blacks was now 83.4% in relation to whites, or within the boundaries of the 80% rule. However, adverse impact was inferred based on: (1) a significant difference on mean scores for blacks (69.17) versus whites (75.59); (2) an effect size of $d = 0.9$ (which is very large); and (3) a test of independent proportions that resulted in a *Z*-test difference of 2.35 standard deviations.

To summarize, five major points were made in this section:

1. The *Guidelines* were written to unify regulations on adverse impact across five federal agencies. They were adopted in 1978.
2. The authors of the *Guidelines* recognized that the guidance provided would become dated over time and, therefore, a provision was included to keep abreast of new validity strategies as, for example, written into the *Principles* and *Standards*.
3. Guidance suggesting that content validity strategies could not be used for mental abilities was struck down in *Guardians v. CSC* (1980), a ruling that subsequently was consistently supported in the courts.
4. The *Connecticut* ruling makes it imperative to evaluate each step in a multiple hurdle procedure both for assessment of adverse impact at each step and to ensure that ratings, particularly subjective ones, are reliably established. Again, if the components of the multiple hurdle procedure cannot be disaggregated, the employer faces the dubious task of having to validate the selection procure as a whole.
5. The $4/5^{th}$ or 80% rule should be treated as a rule of thumb. In general, it is more important to assess statistical significance and effect sizes. Particular attention should be paid to the flip flop rule, particularly with small sample sizes.

1.5 Section 4: A Best Practice Procedure for Job Analysis and Developing and Validating a Text Battery

The author used the following four steps to evaluate a test battery for a large technology company:

- Step 1: A comprehensive job analysis
- Step 2: Development of a test battery
- Step 3: Development of an electronics knowledge test
- Step 4. Criterion validation of the test battery

1.5.1. Step 1: Comprehensive Job Analysis

Job analysis is a prerequisite task for both predictive and content validity strategies. A comprehensive job analysis establishes (a) the essential job tasks (or functions) for performing the work and (b) the knowledge, skills, and abilities (KSAs) needed to perform the essential tasks. The end product of a task analysis is termed **job description**, whereas the end product of an assessment of KSAs is termed **job specification**.

The author favors a method termed CJAM (combination job analysis method) as the best strategy for most purposes, but particularly for validating written tests of cognitive ability. The CJAM strategy was initially proposed by Levine (see Brannick, Levine, & Morgeson, 2007) and has separate strategies for the descriptive and specification phases. The author includes a third component, which is to organize the job tasks into major duty areas.

The following is a summary of actual events surrounding the development and validation of a test battery to hire entry-level technicians that maintain and repair equipment used to manufacture wafers and assemble chips. The case study includes both a content validity and a criterion validity component. Portions of both studies were done contemporaneously. The focus in this section is on how the job analysis was actually conducted.

Prior to the description and specification phases, the author met with the director of human resources and two engineers who served as SMEs. After reading written material describing the entry-level technician job, the author outlined the general steps needed to ensure that all participants understood the mission and were on board. Based on preliminary reading, the author inferred a total of nine major duty areas for the entry-level technician job. The SMEs reported that two of these areas were essentially the same (troubleshooting and equipment repair) and the following eight major duty areas were established:

1. Repetitive tasks
2. Training new technicians
3. Use of test equipment
4. Working with engineering personnel
5. Troubleshooting and repairing equipment
6. Record keeping
7. Ordering spare parts/equipment
8. Safety

Subsequently, two independent SME groups were established, each with six members. The members in both groups consisted of four job incumbents and two supervisors. Both groups (SME1 and SME2) were representative of race, gender, and seniority of entry-level technicians.

In the job task phase, SME1 was presented with the major duty areas and asked to compose critical job task statements within each duty area in a focus group format. Each job task statement was structured to represent an activity, with or without equipment, for a specific purpose. Examples were provided (e.g., inspection of equipment, with meter, to determine proper functioning). Each job task was then independently rated by SME1 members on a 5-point scale for **task complexity** and **consequences of error**. These ratings were then multiplied for

each SME1 member, and means were computed. A mean multiplicative value of 10 was used as benchmark to operationally define critical tasks. SME2 examined and rated the tasks generated by SME1 in a shortened version of the SME1 procedure, and a final list of critical tasks was generated.

In the KSA phase, SME1, operating as a focus group, generated KSAs associated with the critical tasks. The KSAs were then graded by individual members on the basis of four questions:

1. Is the KSA necessary for new workers? *(Answer yes or no.)*
2. Is it practical to expect a new hire to have the KSA? *(Answer yes or no.)*
3. Is trouble likely if the KSA is ignored? *(Rate 1–5.)*
4. Does the KSA distinguish average and superior performers? *(Rate 1–5.)*

As in the job task phase, SME2 verified the answers provided by SME1. A total of 62 KSAs were generated; 18 for knowledge, 10 for skills (with equipment), 29 for general abilities (without equipment), and 5 personal characteristics.

A *yes* answer to Question 1 was required to qualify for the final group of KSAs. A *yes* answer to Questions 1 and 2 meant a KSA would require training prior to employment. Among KSAs with *yes* answers to question 1, mean multiplicative ratings of 10 or higher on Questions 3 and 4 were used as the benchmark, and KSAs above benchmark were examined to determine suitability for testing versus other methods of assessment (e.g., resumes, background checks). The following knowledge (K) areas and abilities (A) were deemed suitable for testing:

Knowledge	**Abilities**
Units of measurement	Problem solving
Electrical systems	Deductive reasoning
Basic math	Reading comprehension
Basic electronics	Following directions
Basic mechanics	Written instructions

1.5.2. Step 2: Development of a Test Battery

There are no hard and fast rules for developing a test battery. The most important consideration is to ensure that all of the KSAs inferred from the job analysis are represented. Beyond that, there are three rules the author has relied on in the past.

The first rule is to use, where possible, that which has been used and validated in-house in the past. This is obviously cost efficient, but it also provides a strong foundation for validity in that it is already supported, it typically results in additional data to supplement the previously established findings. For present purposes, the company had previously created and validated an eight-component test battery of which four of the components were relevant. The KSAs represented by these components were basic math, reading comprehension, following directions, and following written instructions.

The second rule is to study what is available off the shelf. Again, efficiency is a factor here. Judicious use of pre-existing, well-established assessments can save considerable resources in both development and validation efforts. However, one must be very careful to select quality, relevant testing products. The consequence of error here could be expenditure of even more resources than what would have been needed to develop a tailored, in-house solution. That said, there are many readily available, well established tests of many constructs, and as long as the practitioner has sufficient understanding of their use and application, there is no reason they can't be effectively implemented. For present purposes, the author selected three major tests from the Differential Aptitude Test (DAT) battery including a test of abstract reasoning, which assesses problem solving and deductive reasoning, a test of mechanical reasoning, which assesses only mechanical reasoning, and a test of spatial relations, which assesses problem solving and deductive reasoning.

The third rule is to create a test when needed, which is the most difficult and cost-intensive of the tasks. Development and validation of an effective measure is resource intensive and requires extensive understanding of the relevant constructs and test development expertise. It may be worthwhile to consider outside contractors who specialize in test development for this work. Alternatively, in-house knowledge of the constructs combined with sufficient test-development capabilities can yield effective solutions, and the lessons learned during development may provide long-term gains in the form of increased institutional knowledge and broader testing capabilities. To this point, all but three of the requisite KSAs were represented. There was nothing in-house that was applicable and the author determined that there was nothing off the shelf that could cover all three KSAs. Three additional KSAs (not previously sampled) were units of measurement, electrical systems, and basic electronics. Therefore, it was decided to create an electronics knowledge test (EKT) to assess these KSAs. Although not necessary, it was judged that the EKT test, once constructed, also assessed basic math, problem solving, and deducted reasoning. Table 6 depicts how these tests were combined to sample all ten requisite KSAs.

Table 6. Final Test Battery Plan

	In-House Tests	**Abstract Reasoning DAT**	**Mechanical Reasoning DAT**	**Spatial Relations DAT**	**Electronic Knowledge Test**
Units of measurement					X
Electrical Systems					X
Basic Math	X				X
Basic Electronics					X
Basic Mechanics			X		
Problem Solving		X		X	X
Deductive Reasoning		X		X	X
Reading Comprehension	X				
Following Directions	X				
Written instructions	X				

1.5.3. Step 3: Development of the EKT Test

The EKT was developed using a single group of SMEs, all engineers. There were two sessions. In Session 1, the SMEs identified five major electronic knowledge components, including the following:

1. *Ohm's law calculations*: Calculation of voltage, current, and/or resistance based on circuit diagrams
2. *Rectification, regulation, and induction*: Circuit diagrams to examine knowledge of the function of rectifiers, transformers, and capacitors with respect to both inputs and outputs
3. *Component identification*: Recognition of component parts, including capacitors, electrolytic capacitors, crystals, transistors, Zener diodes, inverters, gates, mosfets, and buffer amps
4. *Component knowledge*: How the component's parts are used
5. *Troubleshooting*: Troubleshooting (or problem-solving) questions, including items based on oscilloscope diagrams, nonelectrical tracing circuits, and circuits taken directly from manuals covering the firm's actual equipment

In Session 2, SME1 generated between 5 and 15 items for each knowledge area for a total of 60 items. Subsequently, the SMEs individually rated each item on a 5-point scale, where *1* = low value, *3* = average value, and *5* = high value. Individual test items were examined using correlations between each item and total score, percentage correct per item, and the aforementioned SME item ratings. Five items were eliminated on this basis, yielding a final EKT with 55 items. Internal consistency reliability for the final 55 items was assessed using Cohen's alpha, which was acceptable at .90. The alphas for the five components of the test were also acceptable, ranging from .85 to .92.

1.5.4. Step 4: Criterion Validation of the Test Battery

At this point, the criteria for establishing content validity for the test battery were satisfied. The next step was to conduct a criterion validity study. There are two general problems associated with such a study. First, there must be a large enough sample size (usually 200 or more) and there has to be adequate performance appraisal data. Unfortunately, there were only 80 job incumbents available for testing. More importantly, the in-house performance ratings, which were on a scale of 1 to 5 (poor to excellent) suffered from range restriction as most of the ratings for the 80 incumbents were 3 or 4. This was an anticipated outcome. To solve the problem, an experimental performance appraisal study was conducted.

Raters (all engineers) were gathered for a three-hour training session conducted by the author. The author explained that these were experimental ratings and were for research purposes only and would be held in strict confidence. Thus, they would not affect the status of any of the entry-level technicians appraised. The author explained the statistical problems associated with restriction in range and rater biases (e.g., halo effects, leniency, stringency, central tendency). Incumbents were assessed on four individual dimensions and a global rating as follows:

1. Quality orientation/attention to detail
2. Problem solving/judgment

3. Initiative
4. Technical knowledge
5. Global rating

Primary and secondary ratings were obtained. Primary ratings were for entry-level technicians under direct supervision, and included ratings on a 9-point scale for both the four dimensions and the global rating. Secondary ratings were for entry-level technicians supervised by others but who knew the job incumbents and included only the global ratings.

Raw scores for the in-house test, DAT subtests, and EKT were converted to *z*-scores, and *t*-scores were computed by adding 10 to each individual *z*-score. Thus, by definition, mean performance was 10.0 for the in-house test, each subtest of the DAT, and the EKT.

Preliminary descriptive statistics (means and standard deviations) were then computed for each of the predictors. They were comparable across the in-house test, each of the three DAT subtests, and the electronics knowledge test for the sample as a whole. However, Level II entry-level technicians scored significantly higher on the EKT (but not the in-house test or DAT). Critically, the correlation between primary and secondary global ratings resulted in an acceptable reliability coefficient (r = .73).

In the end, there was strong evidence of criterion validity for the individual tests, incremental validity for the test battery as a whole, and utility for the EKT and a composite of the in-house test and DAT subtests. The individual correlations (i.e., validity coefficients) were moderated by job level. The validity coefficients for Level I entry-level technicians were .59 for the in-house test, .53 for the DAT, and .35 for the electronics knowledge test. The validity coefficients for Level II entry- level technicians were .31 for the in-house test, .27 for the DAT, and .51 for the electronics knowledge test. All validity coefficients were corrected for unreliability, and both the raw and corrected validity coefficients were reported. A multiple regression analysis showed incremental validity. That is, there were significantly higher multiple R values for the test battery as a whole as compared to individual validity coefficients for the whole sample, and for Level I and Level II entry-level technicians.

Utility was estimated by using global performance appraisal ratings of 5 or 6 (out of 9) as an estimate of average performance and global ratings of 7 or higher as an estimate of superior performance. The percentages of the entry-level technicians showing average and superior global ratings were 69.7% and 21.1%, respectively, for the whole sample. In comparison, these percentages were 78.9% and 39.5% for the top one-third performers on the EKT, and 76.3% and 31.6% for the top one-third performers on a composite of the in-house test and DAT subtests.

To summarize, six major points were made in this section:

1. A comprehensive job analysis is a prerequisite task to establish both content-related and predictive validity.
2. A test plan should be established to determine the KSAs that are suitable for testing.
3. A criterion-related validity study for a test battery requires consideration of appropriate in-house tests, off the shelf tests and the need, where necessary, for creating a new in-house test.

4. A decision needs to be made to determine if a performance appraisal study is required.
5. The test then must be evaluated using a criterion-related strategy.
6. The strategy should include evaluation of individual components of the test as well as the battery as a whole.

1.6 Section 5: How and when to employ content validity strategies

In the author's opinion, it is necessary to perform a criterion study in order to validate off-the-shelf tests. The reason is that such tests are copyrighted and items can neither be added nor removed. Therefore, it is difficult to connect such a test directly to the critical KSAs resulting from a job analysis. However, when a job analysis serves as the basis for a test from the start, a properly conducted content validity study will alone suffice to validate the test. This was illustrated in several court cases cited in Section 3 (e.g., *Gillespie v. Wisconsin* (1985); *Police Officers v. City of Columbus* (1990); *Brunet v. City of Columbus* (1993); *Williams v. Ford Motors* (1999); *Association of Mexican-American Educators v. California* (2000); *Bew v. Chicago* (2001); and *Gulino v. New York State* (2006). In fact, the EKT test discussed in the last section would alone suffice to validate the test.

Recall, there are five steps in a properly conducted content validity study:

1. Suitable job analysis
2. Reasonable competence in test construction
3. Test content related to job content
4. Test content representative of job content
5. Scoring systems selecting applicants who are better job performers

To illustrate, the author recently created and content-validated a group of factory jobs he deemed were in the same job family. As a precursor to the job analysis, the author spoke to the director of recruitment and then observed the family of jobs for two hours. He then made a second visit in which he conducted the actual job analysis. The job analysis resulted in the following critical KSAs:

1. lifting
2. bending
3. squatting
4. twisting
5. stretching

The author then created three simulations to test the various KSAs.

Simulation 1: This simulation was designed to test the ability to lift weight from floor level to above head level. The weight of the product was 30 pounds so as to require only moderate exertion. In any given cycle, the applicant would lift the 30-pound product from the floor to the height of a conveyer belt and back down to the floor followed by a brief (five-second) pause. The cycling process continued for three minutes, after which, there was a two-minute rest period before continuing on to Simulation 2.

Simulation 2: This simulation was designed to test the abilities to bend to the waist and below and to do complete squats. Starting from a full upright standing position with the 30-pound product held chest high, a complete cycle consisted of bending to the waste, then continuing to the floor, returning to the start position, and then performing a complete squat and again returning to the start position. Each cycle was followed by a brief (five-second) pause. The cycling process continued for three minutes, after which, there was a two-minute rest period before continuing on to Simulation 3.

Simulation 3: This simulation was designed to test the abilities to twist laterally in both directions and the ability to stretch with arms extended. This simulation was performed with the same 30-pound product used in Simulations 1 and 2. The applicant started with the product at waist high with arms held close to the body. A complete cycle consisted of a full body twist to the left and returning to the start position, a full body twist to the right and returning to the start position, and a full outward stretch at shoulder level with arms fully extended before returning to the start position. Each cycle was followed by a brief (five-second) pause. The cycling process continued for three minutes, after which, testing was complete.

The three simulations were sufficient to satisfy the five critical steps cited above. First, there was a reasonable job analysis. Second, the creator had the expertise to create the simulations. Third, the simulations were connected to KSAs cited in the job analysis. Fourth, the simulations were representative of all of the KSAs required for the jobs. Finally, to make sure that the test was reliably graded, it was decided that two or more raters independently observed and rated applicant performance.

To summarize, two major points were made in this section:

1. Off-the-shelf tests should be criterion validated.
2. A content validity strategy is sufficient when the appropriate steps are followed in creating a test from the start.

1.7 Section 6: Court rulings related to minimizing adverse impact

Few issues have been more controversial over the past three to four decades than reducing adverse impact. Table 7 summarized methods to reduce adverse impact. Attempts to reduce and/or eliminate adverse impact have targeted mainly cognitive tests. Most of the major cases have involved hiring or promotion decisions in municipal police and fire departments. The purpose of this section is to examine how methods of reducing adverse impact have fared legally. The actual pros and cons of implementing these methods is discussed in greater detail in Ployhart (in press).

Table 7. Methods for Reducing Adverse Impact

1. Subgroup Norming	Use of different norms for minority and nonminority groups; outlawed in the race-norming proscription in CRA-91
2. Banding	Arranging test scores in bands as an alternative to strict rank ordering
3. Alternative Tests or Combinations	Use of different tests or a combination of tests to select applicants
4. Manipulating Test Content	Controversial procedure in which components of a test that produce adverse impact are eliminated
5. Discarding Test Results	Discarding an entire test after it was administered and scored

1.7.1. Method 1: Subgroup Norming

In the early 1980s, the U.S. Employment Service (USES) used subgroup norming to eliminate adverse impact in job referrals by requiring lower percentile scores on the General Aptitude Test Battery (GATB) for referral of minority applicants. This method was supported by the National Academy of Sciences and opposed by the DOJ. The debate was abruptly ended by the **race norming** provision in Section 106(1) of CRA-91, which makes it **unlawful** to "adjust the scores of, use different cutoff scores for, or otherwise alter the results of employment-related tests on the basis of race, color, religion, sex or national origin." It should be noted that the race norming provision is neutral with respect to adding points to test scores for veterans of military service. However, the USES subgroup norming procedure is clearly illegal under this provision.

Technically, subgroup norming is legal if **moderation** by group membership can be proven empirically. However, this would require evidence that the statistical correlation between a test and job performance is different for different protected groups, a result rarely found in the literature (see Guion & Highhouse, 2006).

1.7.2. Method 2: Banding

Bands are ranges (or bandwidths) in which test scores are treated as being statistically equal. In a **race-neutral** approach, selections are made randomly from within bands rather than using a strict rank-ordering. This approach was supported in *Chicago Firefighters Local 2 v. City of Chicago* (2001) where white firefighters challenged random selection for promotion for blacks and Hispanics. The 7th Circuit endorsed this random selection process ruling:

> If banding were adopted in order to make lower black scores seem higher, it would indeed be a form of race norming, and therefore forbidden. But it is not race norming per se. In fact it's a universal and normally an unquestioned method of simplifying scoring by eliminating meaningless gradations.

The problem is that random selection within bands without race-conscious considerations is unlikely to significantly reduce adverse impact.

Race-conscious banding was initially proposed by Cascio, Outtz, Zedeck, and Goldstein (1991). Their argument is that race-conscious banding significantly reduces adverse impact with minimal loss in test utility. This viewpoint has been attacked by several authors (e.g., Gottfredson, 1994; Sackett & Harris, 1984; Sackett & Roth, 1991; Sackett & Wilk, 1994; Schmidt, 1991). The psychometric issues in this debate are complex. Undoubtedly, sole use of race-conscious selection within bands would be a violation of the race norming proscription.

However, limited forms of race-conscious banding has been supported in the courts under restricted conditions. The two most important rulings on this issue are by the 2nd circuit in *Bridgeport Guardians, Inc. v. City of Bridgeport* (1991) and the 9th circuit in *Officers for Justice v. Civil Service Commission* (1992).

In the *Bridgeport* case, the 2nd circuit ruled that the city "could have chosen to use banding in order to alleviate the disparate racial effect of the examination without disserving its legitimate interest." However, the critical issue here was that the banding procedure at issue did not use minority preference as the sole basis for decisions; it was one of **nine** secondary criteria for making selections from within the bands. Stated differently, if applicants are considered equal on the first eight secondary criteria, then minority preference can be used as the tie breaker.

The *Officers* case was somewhat different and more complex. San Francisco was under a consent decree to validate its police promotion exam in accordance with the *Uniform Guidelines*. The city's initial proposal was to make 100 promotions in a strict rank order and to use banding with minority preference for 15 additional promotions. The district court rejected this proposal. The city then proposed using race as one of **four** criteria for selection (along with professional conduct, education and training, and experience) for the 15 promotions. The district court accepted this proposal and the 9th circuit affirmed, calling it "a modified proposal along the lines approved in *Bridgeport Guardians, Inc. v. City of Bridgeport*."

Although promising from the perspective of racial diversity, it should be noted that the *Bridgeport Guardians, Inc.* ruling predated CRA-91 and in the Officer's case, although post-CRA-91, San Francisco was under a consent decree to either validate their test or find an alternative selection procedure with less or no adverse impact.

1.7.3. Method 3: Alternative Tests or Combinations

Unlike subgroup norming and race-conscious banding, alternative selection procedures with less or no adverse impact have a strong legal basis. Such procedures were first endorsed by the Supreme Court in *Albemarle v. Moody* (1975), and subsequently adopted in Section 1607(3)(B) of the *Uniform Guidelines*. Accordingly:

> Where two or more selection procedures are available which serve the user's legitimate interest in efficient and trustworthy workman-ship, and which are substantially equally valid for a given purpose, the user should use the procedure which has been demonstrated to have lesser adverse impact.

Two recent district court rulings supporting alternatives with less adverse impact are *Bradley v. City of Lynn* (2006) and *Johnson v. City of Memphis* (2006). In *Bradley*, a written test was the sole basis for hiring entry-level firefighters. The trial judge ruled that there was adverse impact and insufficient evidence of job-relatedness. More importantly for present purposes, the judge also cited two valid alternatives with less adverse impact: (1) a combination of cognitive tests and physical abilities, and (2) a combination of cognitive tests, personality tests, and biodata. The judge ruled that "while none of these approaches alone provides the silver bullet, these other non-cognitive tests operate to reduce the disparate impact of the written cognitive examination."

In *Johnson*, the trial judge ruled that a promotion exam (for police sergeant) *was* valid, but the plaintiffs won on the theory that alternatives were available with less adverse impact. That ruling was based on the prior development of a valid promotion test in 1996 that resulted in less adverse impact relative to a subsequent promotion exam. The judge ruled "it is of considerable significance that the City had achieved a successful promotional program in 1996 and yet failed to build upon that success."

It is unclear how much weight should be placed on the *Bradley* and *Johnson* rulings given that they are district court rulings. However, regardless of the ultimate disposition of these and other cases at the circuit court and Supreme Court level, it is clear that employers must consider alternative selection procedures with less adverse impact as they commission and/or develop tests that are likely to produce adverse impact.

1.7.4. Method 4: Manipulating Test Content

The Second Circuit's ruling in *Hayden v. Nassau County* (1996) is likely the most controversial circuit court ruling on adverse impact. The case originated in 1977 after the DOJ sued Nassau County. Ultimately, the two parties entered into a consent decree in 1982 to construct an exam that produces no adverse impact, or is valid "in accordance with Title VII and the *Uniform Guidelines*." However, exams developed in 1983 and 1987 also resulted in adverse impact (and two new consent decrees). Then, in 1990, the DOJ and Nassau jointly appointed a Technical Design Advisory Committee (TDAC) to develop a new exam.

The TADC developed a 25-component test and administered it to 25,000 candidates. The total score on the exam resulted in severe adverse impact. Consequently, the TDAC attempted to eliminate adverse impact completely, but deemed the end result invalid. In the exam that was ultimately adopted, 16 of the 25 components were eliminated, resulting in less adverse impact. However, the subsequent nine-component test was challenged by 68 unsuccessful candidates alleging that they would have been selected if all 25 components were used. The 2nd circuit ruled that "the intent to remedy the disparate impact of the prior exams is *not* equivalent to an intent to discriminate against non-minority applicants." The court acknowledged that the decision to "redesign the exam" took race into account. However, the court reasoned that the exam was "scored in a wholly race-neutral fashion." Therefore, the ultimate ruling was that the plaintiffs had failed to state a claim under the Equal Protection Clause of the 14th amendment, Title VII, or the race-norming proscription in CRA-91.

It is difficult to assess the meaning of the *Hayden* ruling. Nassau County was the subject of multiple lawsuits over a 20-year period and had the benefit of a court-sanctioned consent decree

between itself and the DOJ. In light of the Supreme Court's subsequent ruling in *Ricci v. Destefano* (2009), to be discussed shortly, there are stronger legal obstacles facing employers who use the *Hayden* solution absent court approval.

1.7.5. Method 5: Discarding Test Results

In *Ricci v. DeStefano* (2009), the New Haven Civil Service Board (CSB) refused to certify (and thus discarded) firefighter promotion exams for captain and lieutenant positions. There were 41 applicants and seven vacancies for captain, and 77 applicants and eight vacancies for lieutenant. The CSB used a rule of three in which any of three highest scoring applicants could be promoted for any given vacancy. As a result, there were nine applicants eligible for captain and 10 applicants eligible for lieutenant under this rule. No blacks and two Hispanics were eligible for promotion to captain, whereas no blacks or Hispanics were eligible for promotion to lieutenant.

The CSB discarded the exams because they believed they would lose an adverse impact lawsuit by the minority applicants. In response, 17 whites and one Hispanic sued and ultimately won at the Supreme Court level on grounds that the CSB's decision to not certify constituted illegal disparate treatment under Title VII.

The district court judge acknowledged that race was a factor in the CSB's decision. Nevertheless, based on the *Hayden* ruling, the judge ruled that the decision to not certify was race neutral. She also ruled that the "intent to remedy disparate impact of the prior exams is not equivalent to an intent to discriminate against non-minority applicants." A three-judge panel of the 2nd circuit then affirmed this ruling in a short per curiam ruling. Following that, a 13-judge panel of the 2nd circuit refused to review the case en banc in a close 7-6 ruling. The Supreme Court then agreed to review the case based on a written opinion by six dissenters who voted to review the case en banc.

The Supreme Court overturned both lower courts and granted summary judgment for the plaintiffs, thus ending the case. The Court ruled that the motive for discarding the exams was racial, implying that the CSB would have certified the results had they been more favorable for minorities. Accordingly:

> Whatever the City's ultimate aim—however well-intentioned or benevolent it might have seemed—the City made its employment decision because of race. The City rejected the test results solely because the higher scoring candidates were white. The question is not whether that conduct was discriminatory but whether there was a strong basis in evidence that the city would lose an adverse impact challenge.

Nevertheless, the Supreme Court felt it was necessary to balance the tension expressed by the CSB between disparate treatment and adverse impact. He ruled that a *certainty* criterion for losing on adverse impact was too harsh. Accordingly:

> Forbidding employers to act unless they know, with certainty, that a practice violates the disparate-impact provision would bring compliance efforts to a near standstill. Even in the limited situations when this restricted standard could be met, employers likely would

hesitate before taking voluntary action for fear of later being proven wrong in the course of litigation and then held to account for disparate treatment.

The Supreme Court rejected CSB's argument, which was that they had a good-faith effort that they would lose on adverse impact. Instead, the Court demanded that CSB had to be certain they would lose on adverse impact.

To summarize, five major points were made in this section:

1. Race norming is illegal under CRA-91.
2. Race-based banding has been found legal under the limited circumstance that it be used as a tie breaker when there are multiple criteria that precede it, in which case, it can serve as a tie breaker.
3. Using alternative selection procedures as a way of reducing or eliminating adverse impact is a legal strategy.
4. Manipulating test content in order to reduce or eliminate adverse impact has been supported in one case. However, it should be noted that the defense was under a consent decree to either create a test that does not produce adverse impact or to validate the test.
5. In order to discard test results because of adverse impact, the employer must be certain it will lose in court. A good faith belief is insufficient.

2.0 CONCLUSIONS

To summarize, there are six sections in this chapter:

Section 1 was an overview of types of discrimination including distinctions between facial discrimination, disparate treatment, pattern or practice of discrimination, and adverse impact. It was noted that because both use statistical disparities, these two forms of discrimination are often confused with each other.

Section 2 overviewed major Supreme Court rulings on adverse impact. It was noted that the early traditions in *Griggs* and *Albemarle* were temporarily altered in *Wards Cove* before being recovered, for the most part, in CRA-91.

Section 3 focused on key guidelines in the *Uniform Guidelines* that were overturned in the court. This included overturning guidance prohibiting use of content validity to validate tests of mental ability and the *Connecticut* ruling relating to multiple hurdles. Thus, adverse impact must be defended if there is any adverse impact in any of the steps.

Section 4 focused on how to create a test battery based on critical KSAs derived from a job analysis. The discussion included previously validated in-house tests, off-the-shelf tests, and newly developed tests. A criterion validity study was illustrated.

Section 5 first made the point that off-the-shelf tests should be criterion validated because they are copyrighted and their items cannot be removed or added to connect to critical KSAs of a job analysis. It was noted that well-conducted content validity studies (from scratch) can be used for cognitive tests. A method for content validating a newly developed test can be used and connected to critical KSAs from a job analysis.

Section 6: The final section focused on methods for reducing adverse impact. It was concluded that race norming is strictly illegal under CRA-91; that a limited form of race-conscious banding may be used as a tie breaker after prior steps have been satisfied; that using alternative test or selection procedure that reduces or eliminates adverse impact is legal; that altering items on a test after it has been administrated has been supported in one case, but this procedure was used because of a consent decree; and lastly, a good faith effort of losing an adverse impact claim is not sufficient - certainty of losing such a claim is the criterion.

REFERENCES

Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975).

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Anastasi, A. (1982). *Psychological testing* (5th ed.). NY: MacMillan.

Association of Mexican-American Educators v. State of California, 231 F.3d 572 (9th Cir. 2000).

Bew v. City of Chicago, 252 F.3d 891 (2001).

Bradley v. City of Lynn, 576 F.2d Supp. 204 (2006).

Brannick, M. T., Levine, E. L., & Moregeson, F. (2007). *Job and Work ANALYSIS: Methods, Research, and Applications for Human Resource Management.* Thousand Oaks, CA: Sage.

Bridgeport Guardians, Inc. v. City of Bridgeport, 933 F.2d 1140 (1991).

Brooks v. AFC Industries, 537 F. Supp. 1122 (1982).

Brunet v. City of Columbus, 1 F.3d 390 (1993).

Cascio, W. F., Outtz, J., Zedeck, S., & Goldstein, I. L. (1991). Statistical implications of six methods of test score use in personnel selection. *Human Performance, 4*, 233–264.

Chicago Firefigthers Local Union 2 v. City of Chicago (CA7 2001) 49 F.3d 649

Connecticut v. Teal, 457 U.S. 440 (1982).

Diaz v. Pan American World Airways Inc., 442 F.2d 385 (1971).

Domingo v. New England Fish, 727 F.2d 1429 (1984).

Dothard v. Rawlinson, 433 U.S. 321 (1977).

Gilbert v. Little Rock, 799 F.2d 1210 (1983).

Gillespie v. Wisconsin, 771 F.2d 1035 (1985).

Gottfredson, L. S. (1994). The science and politics of race norming. *American Psychologist*, *49,* 955–963.

Griggs v. Duke Power, 401 U.S. 424 (1971).

Guardians v. Civil Service, 633 F.2d 232 (1980).

Guion, R. M., & Highhouse, S. (2006). *Essentials of personnel assessment and selection*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Gulino v. New York State, 460 F.3d 361 (2006).

Gutman, A. (2005). Adverse Impact: Judicial, Regulatory, and Statutory Authority. In F.J. Landy (Ed.) *Employment Discrimination Litigation: Behavioral, Quantitative, and Legal Perspectives*. San Francisco, CA: Jossey Bass, pp. 20-46.

Hayden v. Nassau County, 180 F.3d 42 (1996).

Hazelwood School District v. U.S., 433 U.S. 299 (1977).

International Teamsters v. US, 431 U.S. 324 (1977).

Isabel v. City of Memphis, 404 F.3d 404 (2005).

Johnson v. City of Memphis, 355 F.2d Supp. 911 (2006).

Landy, F. J. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist, 41,* 1183–1192.

Landy, F. J., Gutman, A., & Outtz, J.L. (2010). A sampler of legal principles in employment selection. In J. L. Farr & N. T. Tippins (Eds.), *The Handbook of Employee Selection.* New York: Taylor & Francis Group, pp. 229–255.

Levin v. Delta Airlines Inc., 730 F.2d 994 (1984).

McDonnel Douglas v. Green, 411 U.S. 792 (1973).

Morris, S. B., & Lobsenz, R. E. (2000). Significance tests and confidence intervals for the adverse impact ratio. *Personnel Psychology, **53,*** 89–111.

Officers for Justice v. Civil Service Commission, 979 F.2d 721 (1992).

Ployhart, R. (in press). *Air Force Personnel Center Best Practices Guide Selection and Classification Model Development*.

Police Officers v. City of Columbus, 916 F.2d 1092 (1990).

Ricci v. DeStefano, 557 U.S. 557 (2009).

Sackett, P. R., & Roth, L. (1991). A Monte Carlo examination of banding and rank order methods of test scores used in personnel selection. *Human Performance*, *4,* 279–295.

Sackett, P. R., & Wilk, S. L. (1994). Within group norming and other forms of score adjustment in pre-employment testing. *American Psychologist, 49,* 929-954.

Schmidt, F. L. (1991). Why all banding procedures are logically flawed. *Human Performance, 4,* 265–278.

Siskin, B. R., & Trippi, J. (2005). Statistical issues in litigation. In F. J. Landy (Ed.), *Employment discrimination litigation: Behavioral, quantitative, and legal perspectives* (pp. 132–166). San Francisco, CA: Jossey Bass.

Society for Industrial-Organizational Psychology (2018). *Principles for the Validation and Use of Personnel Selection Procedures* (5th ed.).

Torres v. Wisconsin Department of Health and Social Services, 859 F.2d 1523 (1988).

Uniform Guidelines on Personnel Selection Procedures. EEOC (1978) 41 CFR 60-341 CFR 60-3

Wards Cove v. Atonio, 490 U.S. 642 (1989).

Washington v. Davis, 426 U.S. 229 (1976).

Watson v. Ft. Worth Bank, 487 U.S. 977 (1988).

Williams v. Ford Motors, 14 F.3d 1305 (1999).

Wilson v. Southwest, 517 F. Supp 292 (1981).

LIST OF ACRONYMS, ABBREVIATIONS, AND SYMBOLS

%	percent
A	ASVAB Administrative composite
A1	Air Staff
AERA	American Educational Research Association
AF/A1PT	Air Force Testing Policy
AFPC/DP3SP	Air Force Personnel Center Promotions, Evaluations, and Recognition Branch
AFPC/DSYX	Air Force Personnel Center, Strategic Research and Assessment Branch
AFHRL	Air Force Human Resources Laboratory
AFOQT	Air Force Officer Qualifying Test
AFRS	Air Force Recruiting Service
AFS	Air Force Specialty
AFSC	Air Force Specialty Code
AF-WIN	Air Force Work Interest Navigator
APA	American Psychological Association
ASVAB	Armed Services Vocational Aptitude Battery
ATC	Air Traffic Control
ATST	Air Traffic Scenarios Test
BFOQ	Bona Fide Occupational Qualification
CRA	Civil Rights Act
DoD	Department of Defense
CFM	Career Field Manager
CJAM	Combination job analysis method
CSB	Civil Service Board
CSC	Civil Service Commission
DAT	Differential Aptitude Test
DOL	Department of Labor
DOJ	Department of Justice
E	ASVAB Electronics composite
EEOC	Equal Employment Opportunity Commission
EKT	Electronics Knowledge Test

EPQT	Enlisted Pilot Qualifying Test
EDPT	Electronic Data Processing Test
G	ASVAB General composite
GATB	General Aptitude Test Battery
HS	High School
KSA	Knowledge, skills, and abilities
M	ASVAB Mechanical composite
MEPS	Military Entrance Processing Stations
METS	Military Entrance Test Sites
MDPP	Multidimensional pairwise preference
MTT	Multi-Tasking Test
NCME	National Council on Measurement in Education
OFCCP	Office of Contract Compliance Programs
OA	Occupational Analysis
PCSM	Pilot Candidate Selection Method
RPA	Remotely Piloted Aircraft
SDI	Self-Description Inventory
SIOP	Society for Industrial and Organizational Psychology
SME	Subject matter expert
TAPAS	Tailored Adaptive Personality Assessment System
TBAS	Test of Basic Aviation Skills
TDAC	Technical Design Advisory Committee
USES	U. S. Employment Service