

BY THE COMPTROLLER GENERAL

109356

# Report To The Congress

OF THE UNITED STATES



## Federal Employment Examinations: Do They Achieve Equal Opportunity And Merit Principle Goals?

This study focused on the Professional and Administrative Career Examination (PACE) and the Junior Federal Assistant, Accountant-Auditor, and Social Worker examinations.

GAO found that PACE and the Junior Federal Assistant examination screen out black applicants at a much higher rate than whites and that few blacks who pass the tests score high enough for a realistic job opportunity.

The Office of Personnel Management believes that there is a clear relationship between performance on PACE and job performance. Critics of the test say that this relationship has not been adequately proven.

DLG 00925

This report does not conclude whether PACE or the Junior Federal Assistant examination results in "adverse impact," which is a question the courts must answer, but it does make recommendations to the Director, Office of Personnel Management, which should result in better recordkeeping and test validation.



109356

109356  
005264

FPCD-79-46  
MAY 15, 1979



COMPTROLLER GENERAL OF THE UNITED STATES

WASHINGTON, D.C. 20548

B-167015

To the President of the Senate and the  
Speaker of the House of Representatives

*CWO 00001*

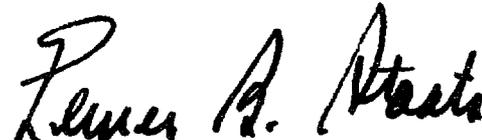
This report discusses a perplexing problem concerning Federal work force development. The goal of Federal employment is to hire qualified individuals on the basis of merit principles, giving everyone an equal opportunity for employment. To achieve this goal the Office of Personnel Management designed various testing and screening devices. The problem is that two major written tests screen out black applicants at a much higher rate than white applicants, making it difficult for agencies to meet their equal employment opportunity goals. This problem is aggravated by the fact that the test which has the greater adverse impact on blacks is also the test which the Office of Personnel Management has researched the most, and which it believes shows a clear relationship between test performance and later job performance.

This report offers no solution to the problem. It does, however, focus attention on the need to renew emphasis on developing selection methods which give everyone an equal chance for employment while assuring a competent and productive Federal work force. The Chairwoman, Subcommittee on Civil Service, House Committee on Post Office and Civil Service, has scheduled hearings for May 15, 1979, to explore these questions.

We obtained formal comments from the Office of Personnel Management, the Equal Employment Opportunity Commission, and the U.S. Commission on Civil Rights. The Department of Justice did not comment because of pending litigation. To the extent time permitted prior to the scheduled hearings, comments were considered in the report.

B-167015

We are sending copies of this report to the Directors, Office of Management and Budget, and Office of Personnel Management; the Attorney General; the Chair, Equal Employment Opportunity Commission; and the Chairman, U.S. Commission on Civil Rights.

  
Comptroller General  
of the United States

COMPTROLLER GENERAL'S  
REPORT TO THE CONGRESS

FEDERAL EMPLOYMENT EXAMINA-  
TIONS: DO THEY ACHIEVE EQUAL  
OPPORTUNITY AND MERIT PRIN-  
CIPLE GOALS?

D I G E S T

*Setup.*

Each year about 1.5 million applications are made for competitive civil service positions, and about 150,000 to 200,000 people are hired from civil service job registers.

By law, selection and advancement should be determined on the basis of ability, knowledge, and skills and after fair and open competition which assures everyone an equal opportunity.

*A query was made to see*  
~~GAO wanted to know~~ whether examining procedures for competitive civil service positions are achieving this U.S. policy goal. *Its* ~~The~~ review focused on the Professional and Administrative Career Examination (PACE), Junior Federal Assistant, and Accountant-Auditor and Social Worker examinations. The last two are evaluations of training, education, and experience listed on a person's employment application and do not require people to assemble at an appointed time and place for a written test.

*FIND/CON.*

GAO found that black applicants passed the written tests at a substantially lower rate than whites and "other" racial minorities, and that few blacks who passed scored high enough to have a realistic chance of employment consideration.

Analysis of PACE scores used for hiring purposes showed that

--58 percent of white applicants passed the test--16 percent with a score of 90 or above, a score high enough for a realistic chance of employment consideration,

FPCD-79-46

Tear Sheet. Upon removal, the report cover date should be noted hereon.

--56 percent of "other" racial minority applicants passed--14 percent with a score of 90 or above, and

--12 percent of the black applicants passed--0.2 percent with a score of 90 or above. (See p. 11.)

Blacks scored somewhat better on the Junior Federal Assistant test than on PACE. However, the proportion of blacks who scored in the certifiable range was still far lower than that of whites--15 percent of whites in the certifiable range, compared to 5 percent of blacks. (See p. 16.)

On the Accountant-Auditor and Social Worker examinations, of those who passed, proportionately as many blacks as whites scored within the certifiable range. (See pp. 18 and 19.)

~~GAO did not reach~~<sup>NO</sup> a conclusion<sup>was reached</sup> concerning the question of adverse impact. The courts ultimately must decide this question, but the Uniform Guidelines on Employee Selection Procedures say that usually the impact of the total selection process for a job will be considered rather than the impact of one component. Other components might include internal promotions and transfers by an agency and special hiring programs where applicants are not required to take a written test. (See p. 20.)

Selection guidelines for Federal agencies have, since 1976, required that records be kept to determine the impact of tests and other selection procedures on specific groups. ~~GAO found that records have not been kept.~~<sup>records approved</sup> to show

--the race or ethnicity of job applicants  
or

--how many individuals, by race, sex, and ethnicity, are selected for an occupation using procedures other than civil service job registers.

As a result, the Office of Personnel Management does not know what impact the selection procedure components have on minorities, nor can it determine whether adverse impact is present based on the total selection process. The Office said that it has been developing a system to collect such data since 1976, but serious funding, forms clearance, and technical problems have hampered implementation. (See p. 21.)

Although records have not been kept on how many individuals entered an occupation using PACE or other entry routes, the Office of Personnel Management estimated that minority persons make up about 17 percent of employees in general schedule grades 5 to 11 in 16 of the most populous occupations for which PACE is used. The Office noted that minority representation in these occupations is higher than relevant labor market statistics, and also higher than minority representation in comparable occupations in the private sector. Although the equal employment opportunity posture in these occupations is relevant, according to the Uniform Guidelines adverse impact is computed on the basis of applicants hired into an occupation rather than on individuals already employed. (See p. 21.)

The Government is obligated to have objective, reliable, valid, and job-related selection procedures. Since GAO's analysis showed that certain tests screened out a disproportionately large number of black applicants, it asked whether a clear relationship between test performance and job performance had been shown. That is, are the tests valid?

The question of test validation has generated much controversy--particularly for PACE, which is used to screen for 118 civil service occupations. The U.S. Civil Rights Commission, the President's Reorganization Project Civil Rights Task Force, and the Department of Justice's Civil Rights Division all have expressed questions about the validity of PACE.

On the other hand, the Chairman of the Civil Service Commission, now the Director of the Office of Personnel Management, believes PACE is the most fully validated and documented test in the Commission's history, and that a clear relationship between job performance and test scores has been shown. The Office of Personnel Management also stated that it has a large staff of experienced selection psychologists while the groups which have criticized PACE lack technical expertise to evaluate validity evidence. (See p. 37.)

The Office of Personnel Management spent a considerable amount of time and money developing PACE and a substantial amount of evidence was gathered to support its validity. (See p. 28 and app. XI.)

GAO reviewed the PACE documentation and considered the frequently voiced criticisms of the validation strategy.

1. Critics have said that the validation strategy used was inappropriate because it is new to the employment field and requires an extensive effort. Although GAO believes the validation strategy was appropriate, standards concerning the need for a representative sample of research participants and investigations of test fairness were not, but should have been, followed. These standards are also included in the Uniform Guidelines which were adopted after the test was developed. In view of the importance of PACE to obtaining an entry level professional or administrative civil service job, and the impact the test has on blacks, GAO believes that the Office of Personnel Management must now scrupulously follow the Uniform Guidelines requirements. (See p. 39.)
2. The test is used for 118 occupations, but detailed job analyses identifying the work behaviors required for successful job performance were conducted for only 27 occupations. GAO believes that

the Office of Personnel Management should assure that the comprehensive approach it planned for performing job analyses of the remaining occupations is completed. (See p. 41.)

3. Criterion-related studies--studies showing a statistical relationship between test scores and measures of job performance--were performed for only 3 of the 12 to 15 occupations where they might have been technically feasible. The Guidelines require a series of criterion-related studies to support the type of validity strategy used, but they do not specify the number of studies needed. GAO believes that the Director, Office of Personnel Management, should determine whether further criterion-related studies are needed. (See p. 43.)
4. The three studies used the concurrent design--using current employees--rather than the predictive design--using applicants--although predictive studies are preferred from a scientific standpoint. GAO believes that, while predictive studies may not be technically possible in a system requiring merit selection, it may be possible to carry out a study where all applicants selected by some other merit procedure are given the test and their job performance is followed up at a later date. (See p. 43.)
5. Critics have pointed out that race and ethnic data on the research participants was not collected or analyzed, and the researchers did not assure that individuals in the study groups were representative of the relevant labor market. GAO believes that if further studies are performed, race and ethnic data should be collected as required by the Guidelines. (See p. 45.)
6. Critics have also pointed out that possible test bias or test unfairness for race, sex, or ethnic subsamples was not

investigated, although this is required by professional standards and the Guidelines. The Office of Personnel Management said that this was not feasible. GAO believes that, if further studies are performed, the Guidelines requirement concerning the investigation of possible test unfairness should be followed. (See p. 45.)

7. Critics have said that suitable alternatives to the test or alternative ways of using the test were not sufficiently investigated. Some alternatives were explored by Office of Personnel Management researchers, but GAO believes more can and should be done to investigate suitable alternatives which have less adverse impact. (See p. 48.)

Documentation to support the validity of the Junior Federal Assistant test used for 28 occupations may not meet current standards. The Office of Personnel Management believes it is valid for all 28 occupations, but it is developing and documenting a new written test to replace it. This project is expected to be completed in October 1982. (See p. 50.)

GAO did not review the Accountant-Auditor and Social Worker examinations to determine their job-relatedness. Instead, it asked what actions had been taken to correct reported problems to make the unassembled examining process more reliable. GAO found that rating guidelines had been improved, a training program had been developed for examiners, and a quality assurance program had been established. In 1977 a new procedure was developed for unassembled examinations. Designed to assess the quality of the applicant's training, education, and work experience, it is being tested on five occupations in five regions during fiscal year 1979. (See p. 51.)

Federal agencies have several methods other than selecting people from a civil service

job register to fill vacancies. The Office of Personnel Management estimates that two-thirds of the administrative and professional career positions are filled each year through employee promotions, transfers, or other staffing procedures not involving a job register. Even for those individuals who have not previously held a Federal job, agencies can use hiring programs which do not require the usual entrance examination. However, these are special purpose programs and the number of positions available is limited. (See p. 59.)

Records are not kept to show how many individuals by race, sex, and ethnicity have been selected for particular occupations using these other selection procedures, and agencies and the Office of Personnel Management do not know whether other individual procedures result in hiring people who perform better or worse on the job than those hired from job registers. (See p. 62.)

#### AGENCY COMMENTS

The Office of Personnel Management, the Equal Employment Opportunity Commission, the Department of Justice, the U.S. Commission on Civil Rights, the Merit System Protection Board, and the Office of Education were given an opportunity to study GAO's preliminary report and comment on the matters discussed. Comments GAO received were considered in the report's final preparation. Justice did not comment because it is currently defending the Government in a suit filed in the District Court for the District of Columbia alleging that PACE is discriminatory and violates the Civil Rights Act of 1964 (Luevano, et al. v. Campbell, No. 79-0271, D.D.C. filed Jan. 29, 1979). The Merit System Protection Board and the Office of Education also did not comment on the report. (See app. XII.)

The Office of Personnel Management said that, in its view, PACE is a fully validated

test for efficiently examining a large number of applicants. It said that the adverse impact of the test is not a phenomenon peculiar to PACE, but is typical of tests used for measuring the kinds of abilities important to success in PACE-type jobs. The search for an alternative of equal validity but less adverse impact is a desirable, but not easily attained, objective. Nevertheless, it has encouraged agencies to use various alternatives to ameliorate adverse impact even without determining whether they were of substantially equal validity.

In response to the frequently voiced criticisms of the PACE validation strategy, the Office of Personnel Management said that neither the Uniform Guidelines nor the Federal Executive Agency Guidelines were in effect when PACE was developed. Further, it said that PACE meets the Uniform Guidelines requirements for interim use of a selection procedure. GAO does not agree that all the standards in effect at the time PACE was developed were met. Those standards called for the collection and analysis of race and ethnic data on research participants, and an investigation of test fairness, which was not done.

The Office of Personnel Management said that job analyses had been performed on all 118 occupations. GAO recognizes that in 1974, as an interim measure, an occupational specialist evaluated the occupations not included in the original research study and made a determination on whether these occupations met the scope of coverage criteria for inclusion under PACE, pending the outcome of a more comprehensive evaluation. GAO could not determine that the comprehensive evaluation described by the Office was completed.

The Office's comments suggested that it believes the criticisms of the PACE validation strategy are not meritorious. Although the Office indicated general agreement with the report recommendations, it made no specific commitment to implement them. However, it is

planning to perform an additional research study of the type suggested in the GAO report. The Office said that it will, to the extent feasible, assure that research participants are representative of the relevant labor market and it will investigate test fairness.

The Equal Employment Opportunity Commission said that the report adequately applies the principles of the Uniform Guidelines to the selection procedures studied. It believes, however, that additional conclusions and recommendations could logically flow from the information in the report.

The Commission said that, on the basis of the information in the report, it would appear that neither PACE nor the Junior Federal Assistant examination meets the Uniform Guidelines requirements for interim use of a selection procedure. The scope of this report did not include a detailed analysis of the interim use question since meeting interim use requirements is not determinate of the validity of an examination, and there is evidence that interim use requirements are being met. The Office has a study underway to meet the Guidelines requirements not already met. However, GAO urges the Commission, in exercising its enforcement function, to independently examine the validity evidence and reach a judgment on the matter of interim use of these written tests.

The U.S. Commission on Civil Rights said that regardless of PACE's validity, few minorities are able to enter Federal employment through this test. It believes the Office should consider replacing PACE with a less discriminatory alternative.

RECOMMENDATIONS TO THE DIRECTOR,  
OFFICE OF PERSONNEL MANAGEMENT

A comprehensive review of the Federal policies and practices affected by the new Uniform Guidelines is being planned by the Office of Personnel Management. This review

will include job qualification standards, examinations, performance appraisals, and promotion policies.

GAO endorses this review. However, actions are needed which cannot await the outcome of the planned review. The Director, Office of Personnel Management, should:

--Increase efforts to comply with the Uniform Guidelines requirement for collecting and maintaining records which will disclose the impact of tests and other selection procedures on the employment opportunities of minority group members and females. As a part of the recordkeeping system required by the Uniform Guidelines, the Office should design a system to track, over a period of time, the job performance of groups of people selected from job registers and those selected using other methods. If the tracking system shows a selection procedure that has less impact on minorities and results in selecting employees who perform as well or better than those selected from the registers, then the use of that procedure should be expanded. If the tracking system shows that employees selected by an alternative method perform less well than those selected from registers, then the alternative should be reexamined and a decision made concerning its continued use.

--Reevaluate the PACE validation strategy to assure that it conforms with the Uniform Guidelines. In the interim, assure that the planned comprehensive job analyses of the remaining occupations are completed.

If further validity studies are necessary, GAO recommends that the Director:

--Reconsider using the predictive design or consider performing a research study in which all applicants selected for a PACE occupation by some other merit procedure are administered the test and followed up at a later date.

--Assure, to the extent possible, that re-  
search participants are representative of  
the relevant labor market.

--Investigate the fairness of the test for  
race, sex, and ethnic subsamples.

C o n t e n t s

	<u>Page</u>
DIGEST	i
CHAPTER	
1 INTRODUCTION	1
Examination and selection	2
Equal employment opportunity	4
Relationship between merit system and the Uniform Guidelines	6
Focus of GAO's study	7
2 HOW WELL DO MINORITIES SCORE ON CIVIL SERVICE EXAMINATIONS?	8
Blacks do not score well on two assembled examinations	9
Minorities score as well as whites on two unassembled examinations	17
Do civil service procedures have adverse impact?	20
Conclusions	24
Recommendations	25
3 DO CIVIL SERVICE TESTS PREDICT JOB PERFORMANCE?	26
What is validation?	26
Has PACE been properly validated?	28
Has the JFA test been adequately validated?	50
Have unassembled examinations been adequately validated?	51
Conclusions	54
Recommendations	55
4 ARE ALTERNATIVE JOB ENTRY METHODS AVAILABLE?	56
Problems in obtaining minorities from registers	56
Alternative staffing procedures	57
Use of alternative methods to achieve EEO objectives	61
Conclusions	63
Recommendations	63
5 SCOPE OF REVIEW	64

	<u>Page</u>
APPENDIX	
I	PACE occupational coverage list 65
II	Department and agency locations contacted 68
III	Prior GAO reports on the subject of EEO in civil service examining and selection 70
IV	Sampling techniques 71
V	PACE - Augmented scores 72
VI	PACE - Raw scores 74
VII	PACE - Outstanding scholars' augmented scores 76
VIII	PACE - Outstanding scholars' raw scores 78
IX	JFA - Junior Federal Assistant examination pass and certifiable rates 80
X	Accountant-Auditor certifiable rates 82
XI	Compendium of publications by the Office of Personnel Management on the development and research of PACE 83
XII	Agency comments 90
	Office of Personnel Management 92
	Equal Employment Opportunity Commission 112
	U.S. Commission on Civil Rights 126
	Department of Justice 128

## ABBREVIATIONS

APA	American Psychological Association
CSC	Civil Service Commission
EEO	equal employment opportunity
EEOC	Equal Employment Opportunity Commission
FSEE	Federal Service Entrance Examination
JFA	Junior Federal Assistant (examination)
GAO	General Accounting Office
GS	general schedule
OPM	Office of Personnel Management
PACE	Professional and Administrative Career Examination
SSA	Social Security Administration

## CHAPTER 1

### INTRODUCTION

The Civil Service Act of 1883 (Ch. 27, 22 Stat. 403) established the Civil Service Commission (CSC) to administer a merit staffing system, including planning and carrying out a competitive examining program, to meet the personnel needs of the Federal competitive service. The act stipulated that:

"The President may prescribe rules which shall provide, as nearly as condition of good administration warrant, for--

"(1) open, competitive examinations for testing applicants for appointment in the competitive service which are practical in character and as far as possible relate to matters that fairly test the relative capacity and fitness of the applicants for the appointment sought."

Competitive examinations are the means by which equality of opportunity is united with efficiency of performance.

At the time of its creation, CSC's principal function was to establish and maintain a merit system and to protect it against advocates of the spoils system. The Civil Service Act, and later the Veterans' Preference Act of 1944, provided procedures under which applicants were appointed to Federal jobs on the basis of their qualifications and fitness, rather than on the basis of personal preference or political considerations.

CSC conducted its operations in headquarters offices in Washington, D.C., and in 10 regional offices and 65 area offices. During fiscal year 1978 it spent about \$35.4 million, or about 30 percent of its operating budget, on recruiting and examining. It processed a total of 1,616,178 Federal job applications and sent 1,108,799 names to Federal agencies, from which 152,771 selections were made.

The Civil Service Reform Act of 1978 (Public Law 95-454), enacted October 13, 1978, reinforced the policy that in order to have a competent, honest, and productive work force reflective of the Nation's diversity, personnel management should be implemented consistent with the merit system principles. One of these principles, as defined by the act, is that:

"Recruitment should be from qualified individuals from appropriate sources in an endeavor to achieve a work force from all segments of society, and selection and advancement should be determined solely on the basis of relative ability, knowledge, and skills, after fair and open competition which assures all receive equal opportunity."

The examination and selection system must be designed to ensure that applicants are fairly examined and evaluated and that their opportunity for employment depends on their qualifications. It must also ensure Federal agencies that the candidates they are considering hiring are well qualified.

#### EXAMINATION AND SELECTION

CSC was responsible for developing and administering an examining program which brought together the principles of equal opportunity and merit in filling competitive jobs in the Federal Government. The new Office of Personnel Management (OPM) assumed these responsibilities on January 1, 1979. According to the Federal Personnel Manual, the merit principles which guide the examining program involve:

- Competitive examinations open to everyone.
- Examination based upon testing and other applicant appraisal procedures which do not discriminate on the basis of race, sex, religion, national origin, or any other factor not directly related to the requirements of the job or jobs involved.
- Equal opportunity for everyone to be examined for positions for which she or he has the minimum qualifications.
- Selection from among the best qualified.

Inherent in a merit system is the need to rate and rank individual applicants on the basis of their fitness for the job. Two different evaluation procedures are used by OPM. The first, and most familiar, is the written test. It is particularly well adapted to the testing of aptitudes or specific knowledge. Written tests, and some performance tests, are referred to as "assembled examinations" because candidates are required to assemble at an appointed time and place to take the test. These tests are used for some clerical and lower level jobs and for the Professional and

Administrative Career Examination (PACE), the recent college graduate's primary means of entry into 118 occupations in the Federal service.

The second procedure used by OPM is an assessment of the candidate's education, training, and experience as shown on his or her formal application. This is called an "unassembled" examination since individuals are not required to assemble. Unassembled examinations are used to evaluate the qualifications of most applicants for trade and labor jobs and mid- and upper-level management jobs.

During fiscal year 1977, approximately 58 percent of all applicants for general schedule (GS) jobs took a written examination or performance test; the remainder were evaluated by unassembled examination procedures. Applicants who receive a high enough score have their scores converted to ratings. Applicants who receive a rating of 70 are considered to have passed the examination and are designated as "eligibles." They are then placed on job registers in order of their scores. If appropriate, 5 or 10 points are added to veterans' scores. As required by law, preference eligibles, such as veterans, are listed ahead of other eligibles receiving the same score. The higher an applicant's score, the better qualified he or she is deemed to be. The higher an applicant is listed on the register, the greater the probability that he or she will be considered for Federal employment. Compensable disabled veterans who pass the test, in addition to receiving 10 extra points, are placed at the top of the register for most jobs at GS-9 and below--even above those with higher scores.

When an agency does not fill a vacancy through promotion or reassignment from within, it requests OPM to provide the names of persons eligible for appointment. Generally, there is a separate register for each job category and grade level. Each register is maintained by one or more OPM offices serving Federal agencies in their geographical areas.

According to law (5 U.S.C. 3318), the agency must select from the three eligibles with the highest scores. This procedure is commonly referred to as the "rule of three." The agency may pass over the three highest eligibles only for legitimate, documented reasons satisfactory to OPM.

## EQUAL EMPLOYMENT OPPORTUNITY

A fundamental merit principle is that employment practices, including examining and selection procedures, must not discriminate on the basis of race, sex, color, religion, or national origin. This principle was incorporated into the Civil Rights Act of 1964 and the Equal Employment Opportunity Act of 1972 which amended the 1964 act to cover Federal employees (42 U.S.C. 2000e-16).

One problem that confronted the Congress that passed the Civil Rights Act of 1964 involved the effect of written preemployment tests on equal employment opportunity. The use of these test scores frequently denied employment to minorities, in many cases without evidence that the tests were related to success on the job. Yet employers--including the Federal Government--wished to continue to use such tests as practical tools to assist in selecting qualified employees. The Congress sought to strike a balance which would prohibit discrimination, but would otherwise permit the use of tests in selecting employees. Therefore, the Congress authorized the use of "any professionally developed ability test provided that such test, its administration or action upon the results is not designed, intended or used to discriminate \* \* \*" (42 U.S.C. 2000e (2)(h)).

The proper use of tests and other examining devices is governed by civil service and EEO laws, court decisions, Federal guidelines on employee selection procedures, and generally accepted professional standards. For several years, Federal EEO enforcement agencies--the Equal Employment Opportunity Commission (EEOC), Departments of Labor and Justice, and CSC 1--had different sets of guidelines for the use of tests and other selection procedures. The guidelines EEOC required private employers to follow were considered to be more stringent than those applied to Federal agency employment practices. On August 25, 1978, after almost 6 years of discussion and drafting of proposed guidelines, the four agencies adopted the Uniform Guidelines on Employee Selection Procedures (43 Fed. Reg. 38290-315; Aug. 25, 1978). The Guidelines require the Federal employer to meet the same standards as private employers.

---

1/All equal opportunity in Federal employment enforcement and related functions vested in CSC pursuant to the Civil Rights Act of 1972 were transferred to EEOC effective Jan. 1, 1979, pursuant to the President's Reorganization Plan No. 1 of 1978.

The fundamental principle underlying the Uniform Guidelines is that employer policies or practices which have an "adverse impact" on the employment opportunity of any race, sex, or ethnic group illegally discriminate unless justified by "business necessity." This normally means that the employer must show a clear relationship between performance on the selection procedure and performance on the job. In the language of industrial psychology, the employer must "validate" the selection procedure. In carrying out the process of validation, the employer must also look for suitable alternative procedures which have less adverse impact than the procedure being used.

The Guidelines define adverse impact as a substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group. EEO enforcement agencies have generally agreed that, for purposes of determining whether to take enforcement action, adverse impact will be determined based upon the total selection process for a job rather than upon the impact of individual components of the process. The Guidelines adopted a "rule of thumb" formula as a practical means of determining adverse impact for use in enforcement proceedings. The formula provides that a selection rate for any race, sex, or ethnic group which is less than 80 percent of the rate at which the most successful group is selected will generally be regarded by the Federal enforcement agencies as evidence of adverse impact.

The courts have final responsibility for interpreting and enforcing the EEO provisions of Federal laws, and the Uniform Guidelines. Generally, the first inquiry that courts make in evaluating employment selection cases is whether selection procedures operate to exclude members of one group at a higher rate than members of another group--a determination of adverse impact. If this is demonstrated, the courts have required evidence showing the job relatedness of the selection procedure which caused the adverse impact. If an employer cannot convince the court that the selection procedure which disproportionately excluded applicants of a race, sex, or ethnic group is justified by business necessity (i.e., job related and no suitable alternative exists), the courts have generally concluded that the procedure illegally discriminates.

There have been few legal challenges to the examining and selection practices used by the Federal Government. One court decision, Douglas v. Hampton, 512 F. 2d 976 (D.C. Cir. 1975), dealt with the racial impact of the Federal Service Entrance Examination (FSEE). The court concluded

that the Congress intended to give public employees the same substantive rights and remedies provided for employees in the private sector. Because FSEE had a racially disparate impact, CSC was required to show that the examination had a demonstrable relationship to successful job performance. While the case was being appealed, CSC discontinued using FSEE and replaced it with PACE. A settlement was reached in the Douglas case, with monetary awards to the plaintiffs.

On January 29, 1979, a case was filed in the District Court for the District of Columbia alleging that PACE has adverse impact on blacks and Hispanics and is discriminatory within the meaning of title VII of the Civil Rights Act of 1964 (Luevano, et al. v. Campbell, No. 79-0271, D.D.C. filed Jan. 29, 1979). This case has not yet been litigated.

RELATIONSHIP BETWEEN MERIT  
SYSTEM AND THE UNIFORM GUIDELINES

The Uniform Guidelines apply to the employment practices of most employers: employers who operate under laws requiring the application of merit system principles--most public employers, including the Federal Government--and private employers who are not required by law to apply merit principles in their employment decisions. The Uniform Guidelines are first and foremost concerned with assuring nondiscrimination. So long as there is no adverse impact, as defined by the Guidelines, an employer may use any selection procedures he wants to use--the procedures may be, but are not required to be, valid or job related.

Merit system principles, however, require that valid selection procedures be used, and, on the basis of this principle alone, the Government is obligated to have valid and job-related selection procedures whether or not adverse impact is demonstrated. If adverse impact cannot be demonstrated, the validation requirements of the Uniform Guidelines do not apply but rather the requirements of the Federal Personnel Manual supplements apply. If adverse impact is found, the Uniform Guidelines require that the employer (1) modify or eliminate the procedure which produces the adverse impact, (2) show a clear relationship between performance on the selection procedure and performance on the job, that is, validate the procedure, or (3) otherwise demonstrate business necessity. In these instances, the Guidelines provide specific technical standards which must be followed in conducting the validation studies.

## FOCUS OF GAO'S STUDY

Federal policies regarding the composition of the civil service work force were stated most recently in the Civil Service Reform Act of 1978 (Public Law 95-454). (See p. 1.) The purpose of our study was to determine whether the procedures for examining people for civil service jobs are achieving the stated policy goals. To do this we first wanted to know whether all applicants for Federal Government jobs--whites and minorities--who take assembled and un-assembled tests pass and receive certifiable scores at the same proportionate rate. We did not review the process by which certified applicants are selected from job registers.

Secondly, if we found that the tests screened out a disproportionately larger number of minority group members, we wanted to determine what had been done to assure that test performance was predictive of job performance. This involved an examination of the documentation prepared by OPM to assure that the tests met both merit system principles and Uniform Guidelines requirements for job relatedness and validity. We also attempted to determine whether agencies have available other employee selection procedures which do not screen out a disproportionately large number of minority group members.

We did not focus on Federal agencies' recruiting efforts or determine whether the Federal Government is as effective as it can be in attracting the most qualified candidates available, nor did we focus on the racial and ethnic composition of the Federal work force to determine how the work force in the occupations covered by the examinations studied compares to the relevant labor market statistics.

## CHAPTER 2

### HOW WELL DO MINORITIES SCORE

#### ON CIVIL SERVICE EXAMINATIONS?

To determine whether civil service examining practices will help Federal agencies achieve the 1978 Reform Act policy goal and whether minorities performed on examinations proportionately as well as whites, we analyzed the results of four examinations--two assembled examinations from which about 12,580 people were hired during fiscal year 1978, and two unassembled examinations from which about 2,180 people were hired.

Our analysis showed that black applicants passed the two assembled examinations at substantially lower rates than white and "other" racial minority applicants. Furthermore, few blacks who passed scored high enough to be considered for employment. Data showed that blacks with the necessary qualifications to apply scored better on the unassembled examinations than they did on the written tests.

Until 1976, CSC's longstanding policy prohibited collection of race, sex, or ethnic group membership data on applicants. However, since then, both the earlier Federal Executive Agency Guidelines applicable to CSC, and the new Uniform Guidelines, require that such records be maintained. The Uniform Guidelines state that:

"Each user should maintain and have available for inspection records or other information which will disclose the impact which its tests and other selection procedures have upon employment opportunities of persons by identifiable race, sex, or ethnic group \* \* \*." (section 4A)

Although this requirement has been in effect since 1976, CSC had not developed a system for collecting data on all applicants nor had it performed a complete analysis to determine the specific impact its examining and selection procedures have on individual groups.

In the absence of job applicant racial data, we asked the Social Security Administration (SSA) to help us develop this information by using both civil service and SSA records. Names and social security numbers of applicants who took selected civil service examinations during fiscal years 1976 and 1977 were matched against the racial identification data in the SSA master data file. SSA's file

contained information on whether the person had identified himself or herself by race as (1) white, (2) Negro, or (3) other/unspecified, or whether the person failed to provide racial identification. Preliminary SSA studies indicate that over 90 percent of those who do not provide racial identification information are white. Both GAO and SSA insured the privacy of the individuals included in our analysis.

On the basis of the information obtained through the SSA master data file, we calculated the pass rates of minorities as compared to whites for two assembled examinations--PACE and the Junior Federal Assistant (JFA) examination. We also compared the performance of whites and minorities on two unassembled examinations--the Accountant-Auditor and Social Worker examinations. In making our analysis for some examinations, we used the entire population of individuals participating, but for other examinations we used generally accepted sampling techniques. (See app. IV.)

BLACKS DO NOT SCORE WELL ON  
TWO ASSEMBLED EXAMINATIONS

Professional and Administrative  
Career Examination

PACE provides agencies with job candidates who demonstrate potential for advancement into professional and administrative positions. It incorporates both a review of education and experience and a written test which all applicants must take to establish eligibility. PACE covers GS-5 and -7 entry level positions in 118 occupational series. These positions involve regulatory and compliance work, administrative and management functions, claims and benefit examining, investigative and law enforcement duties, social services, and other nonclerical work. (See app. I for a list of occupations.) Normal career progression in these jobs is to a GS-9 or higher grade level.

PACE was designed to measure the following abilities which, according to OPM, are critical to successful performance in those occupations:

- Ability to understand and interpret complex reading material, and to use language where precise correspondence of words and concepts makes effective oral and written communication possible.

- Ability to make decisions or take action in the absence of complete information, and to solve problems by inferring missing facts or events to arrive at the most logical conclusion.
- Ability to discover underlying relations or analogies among specific data where solving problems involves formation and testing of hypotheses.
- Ability to discover implication of facts, to reason from general principles to specific situations, as in developing plans and procedures.
- Ability to perform arithmetic operations and solve quantitative problems where the proper approach is not specified.

To pass the written examinations, PACE applicants must score 70 or above and have appropriate experience or a baccalaureate degree to be eligible for employment consideration. As a practical matter, an applicant must have a score of 90 or above to have a realistic chance of having his or her name sent to an agency for employment consideration. During fiscal year 1978, about 135,403 individuals took the PACE, 76,907 passed with a score of 70 or above, and 7,587 were hired from PACE registers.

To analyze how well minorities scored on PACE in relation to whites, we sampled from a universe of 183,468 competitors who (1) took the examination in 1976, (2) had the required education or experience, (3) indicated a willingness to accept a GS-5 starting salary, and (4) had a valid social security number. About 8 percent of those in our universe had failed to provide racial identification. We estimated that, of the applicants with race identification in SSA records, 87 percent were white, 11 percent were black, and 2 percent were "other" races. We made separate analyses of how well different racial groups scored using (1) augmented scores--raw scores augmented by veterans' preference points and/or outstanding scholar provisions, (2) raw scores without any augmentation, and (3) scores for outstanding scholars--individuals with a 3.50 grade point average or in the upper 10 percent of their graduating class. We used separate statistical samples for each universe analyzed. (See app. V through VIII.) At the time we selected our samples, 1976 data was the most recent complete data available.

Augmented scores

Performance on the PACE written test determines the raw score, which can vary between 40 and 100. Raw scores are then augmented, if the applicant is eligible, for veterans' preference and/or outstanding scholar status. Outstanding scholars have their raw score averaged with 100 to produce a final score. According to the scoring system, outstanding scholars cannot receive a final rating below 70. Individuals who pass the examination and meet the Veterans' Preference Act criteria have 5 or 10 points, as applicable, added to their scores.

Our analysis of the augmented scores for those who took the examination nationwide showed that

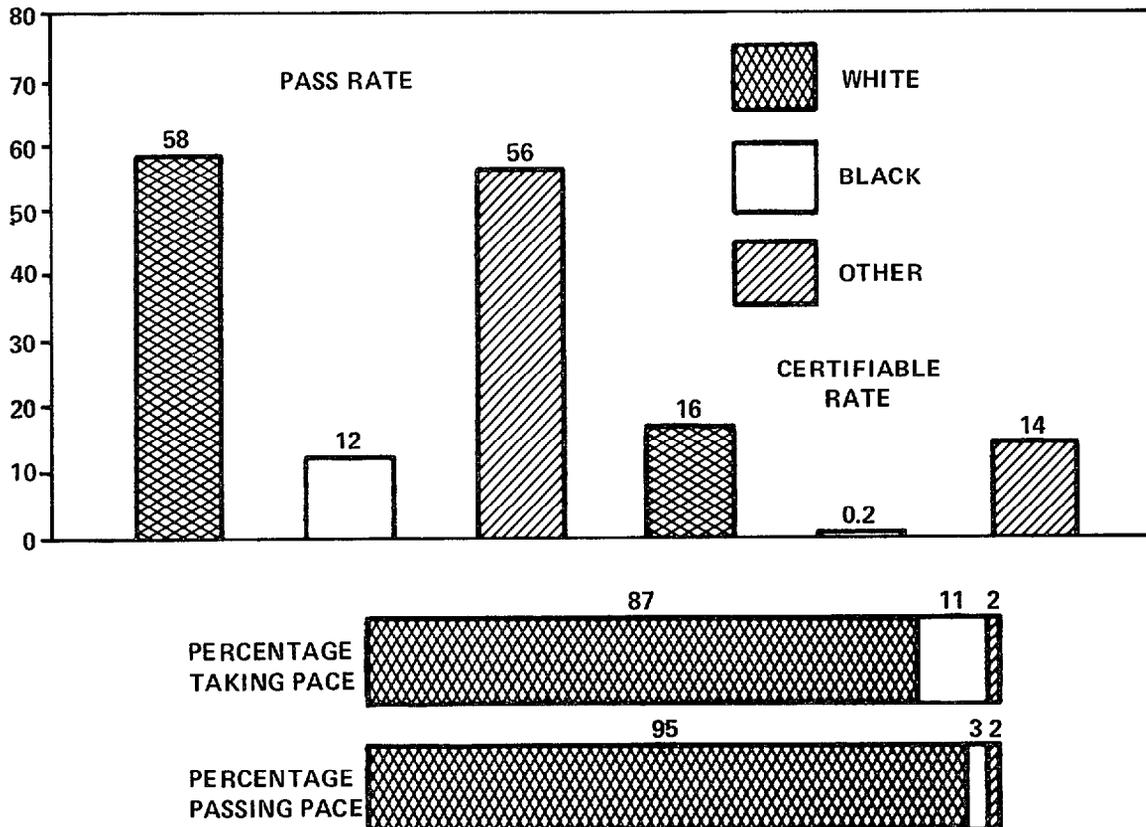
- 58 percent of the whites passed,
- 12 percent of the blacks passed, and
- 56 percent of the "other" races passed.

Of all those with an identifiable race who passed, 95 percent were white, 3 percent were black, and 2 percent were "other" races.

Since a score of at least 90 is needed to have a realistic chance of being certified to an agency for employment consideration, we determined whether blacks were scoring in the certifiable range at the same proportion as whites. We found that they were not. Our analysis showed that 16 percent of whites and 14 percent of "other" races scored high enough to be certified, but less than 1 percent of blacks received an augmented score of 90 or above.

Complete tables showing the results of our analysis in each of the four CSC regions included in our review and all other CSC regions combined are shown in appendix V. The following table and graphs show the result of our nationwide analysis based on a statistical sample of applicants who provided racial identification.

<u>Augmented score</u>	<u>Number of persons</u>		
	<u>White</u>	<u>Black</u>	<u>Other</u>
90 and above	22,583	43	447
70-89	61,276	2,224	1,318
Below 70	<u>61,753</u>	<u>16,575</u>	<u>1,399</u>
Total	<u>145,612</u>	<u>18,842</u>	<u>3,164</u>



Raw scores

To determine whether minorities were scoring on the written test proportionately as well as whites, we analyzed raw scores without considering veterans' preference or outstanding scholar status. Our analysis showed that blacks were scoring disproportionately lower than whites.

--51 percent of the whites who took the written test passed--9 percent with a score of 90 or above.

--4 percent of the blacks who took the written test passed--0.6 percent with a score of 90 or above.

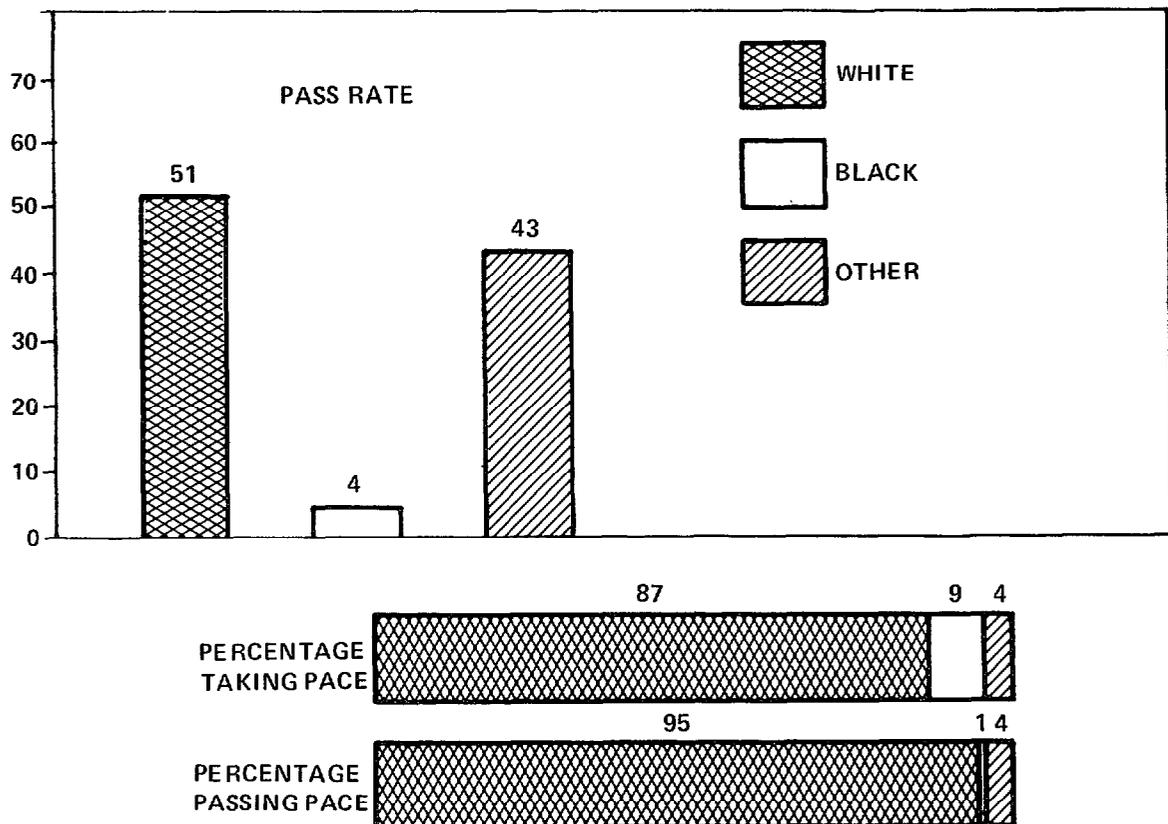
--43 percent of "other" races who took the written test passed--4 percent with a score of 90 or above.

Of those who passed the written examination, 95 percent were white, 1 percent were black, and 4 percent were "other" races.

Complete tables showing results of our analysis in four CSC regions and all other CSC regions combined are in

appendix VI. The following table and graph show results of our nationwide analysis based on a statistical sample of applicants who provided racial identification.

Raw score	Number of persons		
	White	Black	Other
90 and above	13,681	90	298
70-89	60,660	611	2,589
Below 70	72,199	15,071	3,774
Total	146,540	15,772	6,661



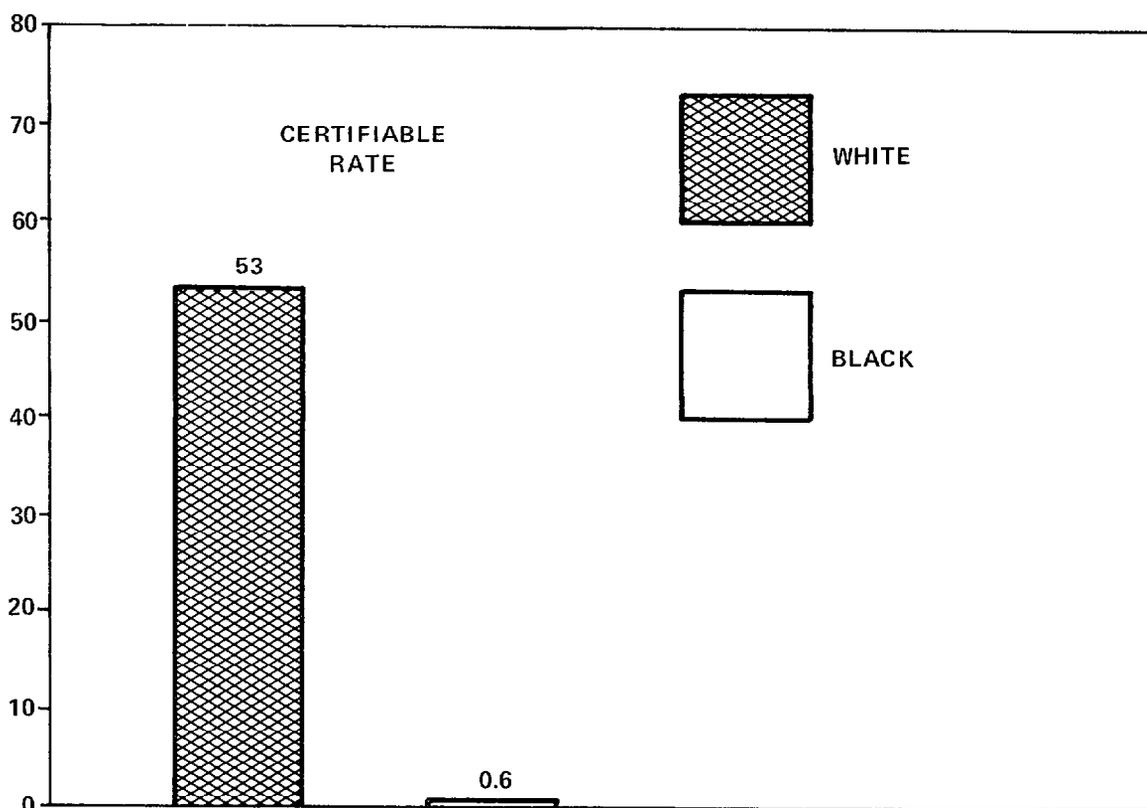
Outstanding scholars

To analyze how well outstanding scholars performed, we sampled from the 26,305 outstanding scholars who took the examination. Outstanding scholars with an identifiable race were 89 percent white and 9 percent black. As previously stated, outstanding scholars, by definition, pass the examination. Therefore, we determined whether black outstanding scholars were scoring in a certifiable range--90 or above--in the same proportions as white outstanding scholars. We analyzed both augmented and raw scores.

Our augmented score analyses showed that blacks were not performing proportionately as well as whites; 53 percent of whites, but only 0.6 percent of blacks, received a score of 90 or above.

Complete tables showing results of our analysis of outstanding scholars' augmented scores in each of the four regions and all other CSC regions combined are in appendix VII. The following table and graph show the results of our nationwide analysis based on a statistical sample of applicants who provided racial identification.

Augmented score	Number of persons	
	White	Black
90 and above	10,994	14
70-89	<u>9,881</u>	<u>2,147</u>
Total	<u><u>20,875</u></u>	<u><u>2,161</u></u>



Our analysis of raw test scores--excluding veterans' preference and outstanding scholar provisions--also showed that blacks were scoring disproportionately lower than whites:

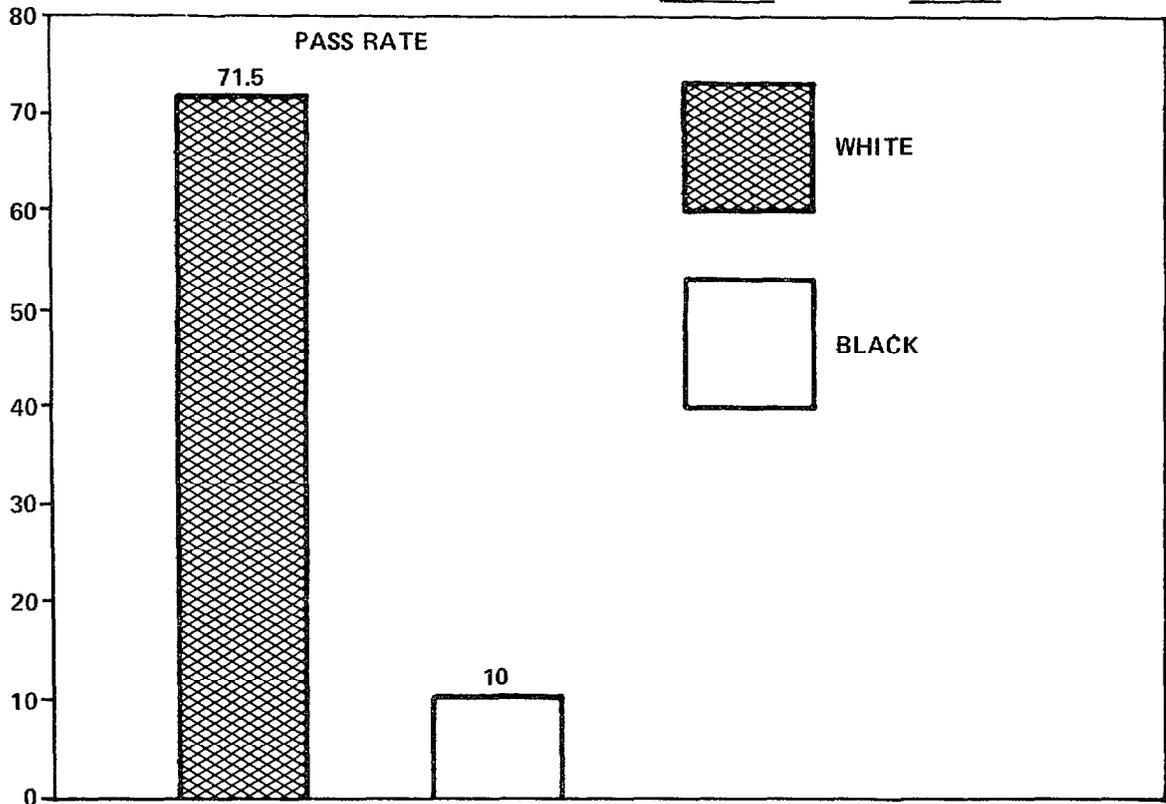
--71.5 percent of white outstanding scholars received a passing raw score--19 percent with a raw score of 90 or above.

--10 percent of black outstanding scholars received a passing raw score--2 percent with a raw score of 90 or above.

Whites accounted for 90 percent of the outstanding scholars sitting for the examination, and blacks accounted for 8 percent. Of the outstanding scholars who received a passing raw score, 98 percent were white and 1 percent were black.

Tables showing the complete results of our analysis in four CSC regions and all other CSC regions combined are in appendix VIII. The following table and graph show the results of our nationwide analysis based on a statistical sample of applicants who provide racial identification.

<u>Raw score</u>	<u>Number of persons</u>	
	<u>White</u>	<u>Black</u>
90 and above	4,187	38
70-89	11,288	154
Below 70	<u>6,173</u>	<u>1,694</u>
Total	<u>21,648</u>	<u>1,886</u>



## Junior Federal Assistant Examination

JFA is used to fill a wide range of technical and administrative support positions. It is designed to measure clerical ability, vocabulary and reading comprehension, and abstract and arithmetic reasoning. It provides entry level employment at the GS-4 level with promotion potential to GS-9.

To be eligible for positions covered by JFA, applicants must pass a written test with a score of 70 and have a minimum of 2 years of college or 2 years of progressively responsible, job-related experience, or a combination of both. The applicant's rating is based on the written test score and an evaluation of education and experience. During fiscal year 1978, CSC processed about 55,300 applications for JFA-type jobs, with about 5,000 individuals hired.

To determine whether minorities scored proportionately as well as whites on the JFA written examination, we analyzed the results of examinations given in four CSC regions during the period July through October 1976. In Atlanta, Philadelphia, and San Francisco, we used for our analysis the universe of those who took the examination. In Dallas we used a statistical sample.

Our analysis showed that blacks scored disproportionately lower than whites in the four regions, as summarized below:

--75 percent of the whites who took the examination passed--15 percent with scores high enough to be certified for employment consideration. 1/

--31 percent of the blacks who took the examination passed--5 percent with scores high enough to be certified for employment consideration. 1/

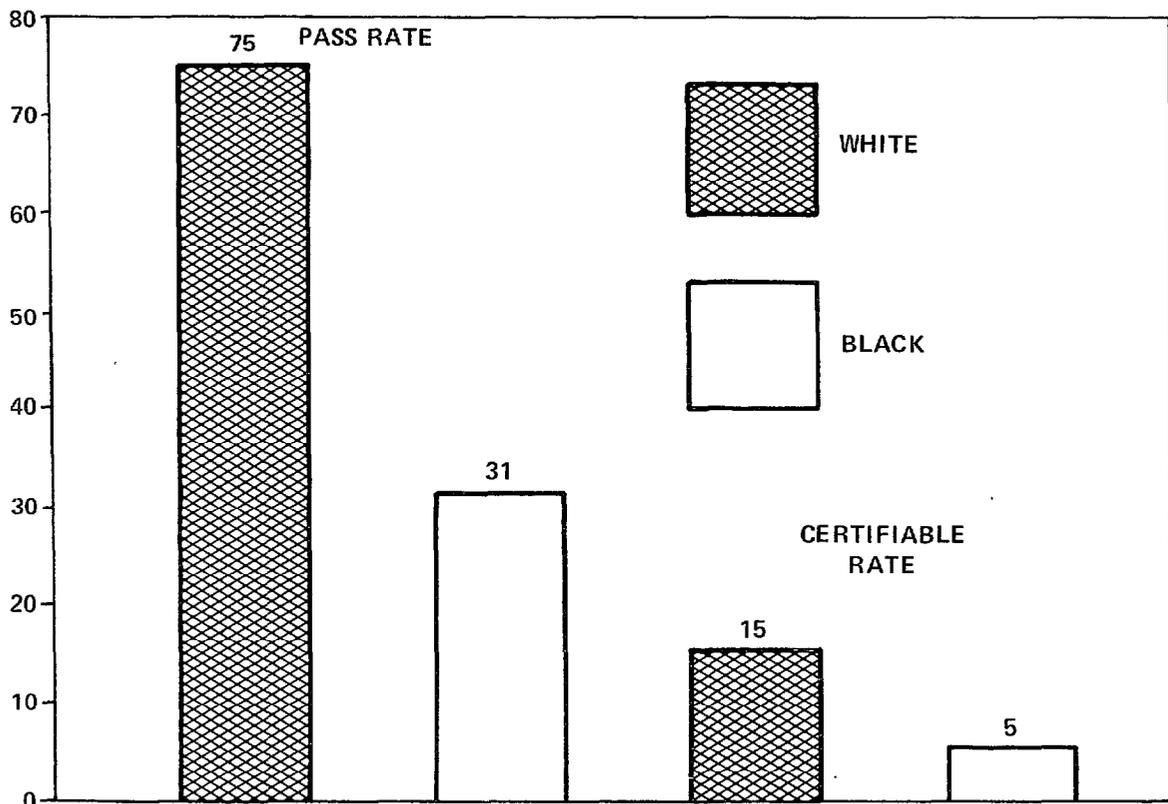
In the four regions, 56 percent of those who took the written test were white and 41 percent were black. The remainder were classified as "other" races.

---

1/Certifiable scores vary from time to time and from region to region depending upon job market conditions. At the time of our review certifiable scores ranged from 70 in the Atlanta region to 90 in the Dallas region.

Tables showing the complete results of our analysis in four CSC regions are in appendix IX. The following table and graph show combined results for the four regions.

	<u>Number of persons</u>	
	<u>White</u>	<u>Black</u>
Certifiable score	158	35
Pass scores below certifiable range	634	204
Failed	<u>268</u>	<u>527</u>
Total	<u>1,060</u>	<u>766</u>



MINORITIES SCORE AS WELL AS WHITES  
ON TWO UNASSEMBLED EXAMINATIONS

Unassembled examination procedures have not been criticized to the same extent as written examinations for denying employment opportunities to minorities. Because the type of data available for assembled examinations was not available for unassembled examinations, we could not determine whether the established qualification standards screened out a disproportionate number of minority applicants.

Nevertheless, for those who met the standards, we determined how well minorities perform in relation to whites on two unassembled examinations--the Accountant-Auditor and Social Worker examinations.

#### Accountant-Auditor examination

The Accountant-Auditor examination is used to evaluate an applicant's ability to function successfully in these positions. It consists of an evaluation of the applicant's education and experience as shown on the individual's application form. Entry level employment is at the GS-5 through GS-9 levels, with promotion potential to higher grades. Applicants must have a bachelor's degree and 24 semester hours in accounting or equivalent experience.

Those who meet the basic qualification standards are considered to have passed the examination and are eligible for employment. Points are added to the basic qualification points for additional education, training, experience, and veterans' preference to determine the applicant's final rating. Applicants for GS-7 positions must also have had 1 year of graduate study or equivalent experience or have had superior collegiate undergraduate achievements--defined as any one of the following: (1) a grade point average of 2.9 on a 4.0 scale, (2) a grade point average of 3.5 in accounting subjects, (3) a standing in the upper third of the class, or (4) election to membership in a national scholastic honor society.

Competition for accountant/auditor jobs is extremely keen. During fiscal year 1978, 19,449 applications were processed for GS-5/7 positions, and 1,916 selections were made.

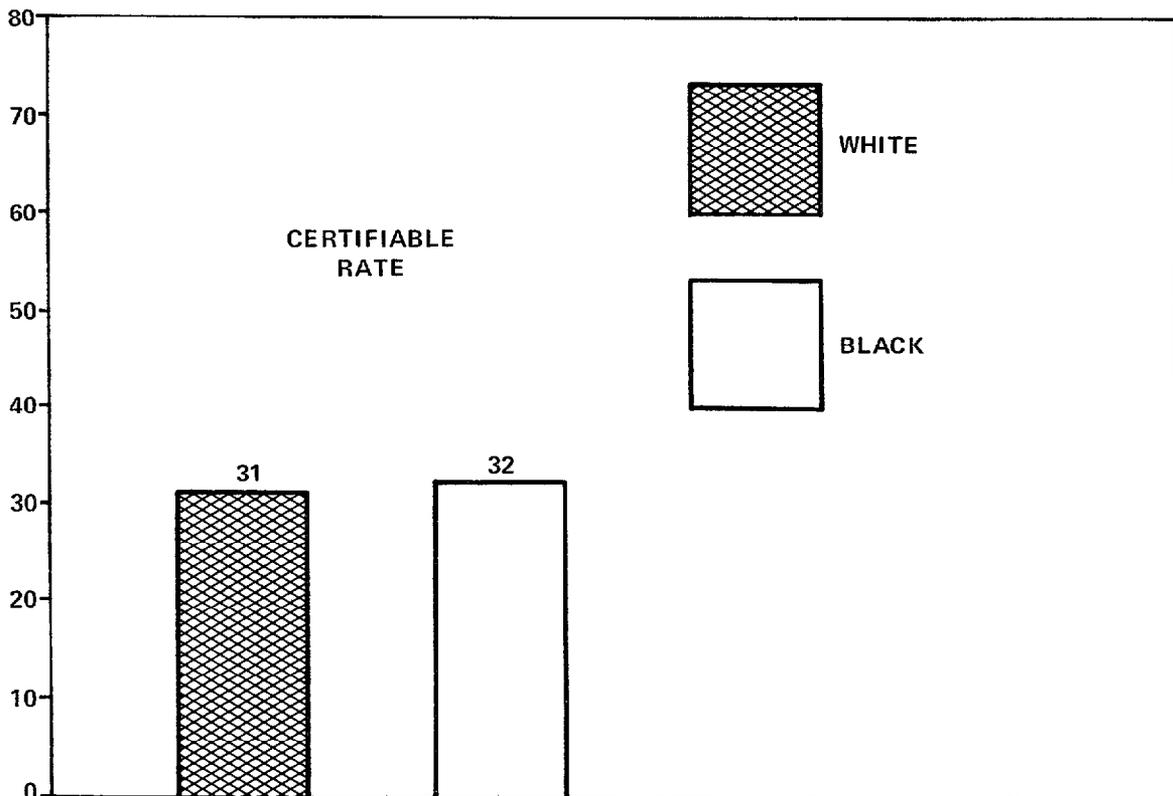
Because all those who meet the minimum qualification standards are considered to have passed the examination, we calculated the rating a person would need to have a realistic chance to be certified to an agency for employment consideration from the GS-7 Accountant-Auditor register. Using this rating, we compared the number of certifiable whites and blacks to the total number of whites and blacks on the register. We made this comparison for all individuals on the registers in Atlanta, Dallas, and Philadelphia and for a statistical sample of individuals on the register in San Francisco.

This comparison showed that, of those who met the minimum qualification standards, blacks were rated equally as high as whites. Data from the four regions showed that

31 percent of whites and 32 percent of blacks were rated high enough to be certified. However, we noted that blacks comprised only 5 percent of the individuals on the registers.

Tables showing results of our analysis in four regions are shown in appendix X. The following table and graph show combined results for the four regions.

	<u>Number of persons</u>	
	<u>White</u>	<u>Black</u>
Certifiable	717	39
Not certifiable	<u>1,613</u>	<u>85</u>
Total	<u>2,330</u>	<u>124</u>



Social Worker examination

This examination is used to evaluate an applicant's qualifications for particular social work positions. The examination consists of an evaluation of the applicant's professional experience, education, and training, and provides entry level employment at grades GS-9 through GS-12. Applicants must have a master's degree in social work.

All applicants who meet basic qualification standards are considered to have passed the examination and are eligible for employment. The names of all eligible applicants are maintained on a referral list centrally in Washington, D.C. A numerical rating is not assigned at the time of application. When an agency requests a list of names to fill a vacancy, applicants' education, training, and experience are evaluated against the requirements of the specific vacancy. Applicants are then ranked, and the names of the top-ranked eligibles are referred to the agency for employment consideration. During fiscal year 1978, 4,048 social worker applications were processed, and 266 selections were made.

We determined how well individuals performed on this examination by sampling from those with an identifiable race who were on the social worker referral list in December 1976. As shown in the following table, blacks were represented on the list of eligibles at a higher proportion than their representation in the U.S. population as a whole--about 11 percent.

	<u>Number of persons</u>	<u>Percent</u>
White	2,119	73.7
Black	681	23.7
"Other" races	<u>76</u>	<u>2.6</u>
Total	<u>2,876</u>	<u>100.0</u>

DO CIVIL SERVICE PROCEDURES  
HAVE ADVERSE IMPACT?

Because of inadequate recordkeeping by OPM and the agencies, we do not know whether civil service employment procedures result in adverse impact which would prompt enforcement action as described by the Uniform Guidelines. Some groups have argued that as long as the "bottom line"--the overall selection process--shows no overall adverse impact, there is no violation of title VII, regardless of the operation of a particular component of the selection process. Others have argued that adverse impact can only be determined by examining each component of the selection procedure, regardless of the bottom line. This question has not been answered definitively by the courts, and the Uniform Guidelines do not address the underlying question of law. They are concerned with how the enforcement agencies will use their resources to combat discrimination. Accordingly, the Guidelines state that:

"If \* \* \* the total selection process [for a job] does not have an adverse impact, the Federal enforcement agencies, in the exercise of their administrative and prosecutorial discretion, in usual circumstances, will not expect a user to evaluate the individual components, and will not take enforcement action based upon adverse impact of any component of that process, including the separate parts of a multipart selection procedure or any separate procedure that is used as an alternative method of selection." (Section 4C)

The Guidelines, however, also state that in unusual circumstances the Federal enforcement agencies may request a test user to evaluate the individual components of the selection procedure for adverse impact and may, where appropriate, take enforcement action with respect to the individual component (section 4C). Most court cases have challenged only one component of a selection process. The bottom line concept, as described in the Guidelines, has not been in force long enough to have a definitive judicial interpretation.

OPM has estimated that only 35 percent of the individuals currently in PACE occupations entered as a result of taking a written test. The remainder entered under other internal and external selection methods, that is, promotion, lateral assignments, cooperative education programs, and other methods. Consequently, they question whether adverse impact, using the bottom line concept, exists. According to OPM's estimates, minority group members comprise 17 percent of the incumbents in grades GS-5 through GS-11--entry level through the most common full performance grade--in 16 populous professional, administrative, and technical occupations for which PACE is used. OPM notes that minority representation in these occupations is higher compared to relevant work force statistics and to private employer representation than it is to relative population statistics.

According to the Uniform Guidelines, the impact that an employer's selection procedures have on minority groups is not determined on the basis of the percentage of minorities currently employed, although the employer's equal employment opportunity posture with respect to jobs or groups of jobs will be considered by the Federal enforcement agencies. Rather, adverse impact is determined on the basis of the rate of selection from the applicant population. To make such a determination, the Guidelines require that:

"Each user should maintain and have available for inspection records or other information which will disclose the impact which its tests and other selection procedures have on employment opportunities of persons by identifiable race, sex, or ethnic group \* \* \* in order to determine compliance with these guidelines."  
(Section 4A)

OPM has not maintained records to show (1) the race or ethnicity of job applicants, (2) how many individuals are selected into an occupation using alternative selection procedures, or (3) the race, sex, or ethnicity of individuals who enter an occupation through alternative selection methods. Until a system is developed to obtain and maintain such information as required by the Uniform Guidelines, the impact which civil service tests and other selection procedures have on the employment opportunities of minorities will not be known.

The question of whether the various selection procedures have an adverse impact on minorities must be answered by the courts or by the administrative agency with EEO enforcement authority. Nevertheless, beyond any EEO requirement, the Government is, in our opinion, obligated to have tests and other applicant appraisal procedures which are valid and job related.

In commenting on this report, OPM said that since 1976, when its policy prohibiting the collection of race and ethnic data was revised, it has been conducting experimental studies on how to collect such data. OPM said that it has a program ready to put a data collection system into operation, but it cited the following problem areas which need to be resolved before the system can be implemented:

- Race and ethnic data collection is extremely expensive, and funds have not been made available to set up a system which would collect the data needed to satisfy the Uniform Guidelines requirements. OPM said that "the cost would run into the tens of millions of dollars."
- The system OPM has devised for data collection is effective when written tests are used, but an effective system has not been designed to collect such data on unassembled examinations.
- The Office of Management and Budget has not authorized the use of forms needed to collect race and ethnic data.

OPM said that once these problems are resolved, it will put in place the procedures to collect data.

In commenting on the magnitude of the adverse impact shown by our analysis, OPM said:

"\* \* \* it is important to recognize that the GAO survey findings may for several reasons exaggerate racial differences in PACE scores:

"a. Competition from private employers for the most able black college graduates, including offers of premium salaries, may have resulted in an underrepresentation of this vital group among those taking the PACE. If this group typically does not take the PACE, this fact would account for part of the test score difference in the GAO data.

"b. Our research indicates that people obtaining lower scores on PACE are more likely to re-take the examination quickly. Since blacks average lower scores as a group, it is possible that proportionately more blacks--especially those with lower scores--have been PACE repeaters than in the case of whites.

"c. The four regions selected for study are not representative of the nationwide PACE population. The two southern regions of Dallas and Atlanta consistently show the lowest average PACE scores. [Our analysis was based on nationwide statistics, not on statistics from four regions.]

"In this connection, the report should also take note of the extensive evidence that adverse impact is not unique to PACE, i.e., the fact that it has been the usual finding among college graduate populations similar to PACE competitors. It occurs, for example, in tests used to determine qualification for entry into professional schools. This fact is essential to an accurate understanding of Federal selection problems because it makes clear that tests like PACE do not create adverse impact. The adverse impact exists long before candidates take the test; the test merely reveals previous real educational deficiencies." (OPM underscoring)

## CONCLUSIONS

Our analysis of available data for two widely used written tests--PACE and JFA--indicated that blacks who took the tests passed at substantially lower rates than whites. Furthermore, few blacks who passed the written tests received high enough scores to have a realistic chance of being certified to an agency for employment consideration.

Our analysis of the two unassembled examinations--Accountant-Auditor and Social Worker--indicated that of those who have the basic qualifications to apply, blacks score proportionately as well as whites on these examinations. Although about the same proportion of blacks and whites on the Accountant-Auditor register were within a certifiable range, we found that only about 5 percent of the individuals on the registers were blacks.

Data presented in this chapter concerning the disproportionately lower pass rates of blacks than whites on the written tests raises questions about their use as preemployment screening and ranking devices in light of the goals set forth in the 1978 Civil Service Reform Act and the Uniform Guidelines on Employee Selection Procedures.

OPM does not currently have information available to show whether individual components of a selection procedure for an occupation, such as PACE or JFA, have an adverse impact on minority group members, nor does it have information to show whether adverse impact is present based on the bottom line concept. The Guidelines recognize that the ultimate determination as to whether an employment practice illegally discriminates is a judicial decision. For administrative purposes in determining when enforcement action will be taken against an employer, the Guidelines state the the bottom line concept will generally prevail. However, the Guidelines do not rule out enforcement action with respect to a specific component of the selection process under unusual circumstances, nor has the bottom line concept been tested in court.

While our analysis indicates that PACE and JFA screen out a disproportionately large number of black applicants, our analysis did not extend to determining specifically who was selected from registers developed on the basis of test results. Until a system is established to obtain and maintain the records required by the Uniform Guidelines, it will not be known whether one component of a selection procedure or the total selection process for a job has adverse impact.

## RECOMMENDATIONS

We recommend that the Director, OPM, act immediately by increasing efforts to comply with section 4 of the Uniform Guidelines on Employee Selection Procedures. This section requires the maintenance of records which will disclose the impact which tests and other selection procedures have on employment opportunities of specific minorities and sex groups.

In addition to the broader requirement of the Uniform Guidelines, we recommend that the Director act immediately to collect data needed to determine the impact on minority groups of PACE, JFA, and other written tests covering more than one occupation, without regard to the bottom line concept. The system used to collect this data should permit a breakdown for each occupation covered by the examinations by OPM region, and allow for a nationwide analysis of the impact of tests on minorities.

CHAPTER 3  
DO CIVIL SERVICE TESTS PREDICT  
JOB PERFORMANCE? .

The purpose of preemployment tests and applicant appraisal procedures is to produce information about the probability of an applicant's job success. The Federal Government's policy and practice has been to use tests and other appraisal procedures that reasonably relate to job requirements so that employment by merit principles and the concept of equal employment opportunity are served and affirmed. This is not an easy task. The methods used to assure that preemployment testing accomplishes this purpose can be extremely complex and sometimes controversial. The complex and controversial nature of the issues involved are exemplified by the nearly 6 years of discussions and negotiations required before the various Federal EEO enforcement agencies could reach agreement on Uniform Guidelines on Employee Selection Procedures. 1/

Because we found that black job applicants passed written tests at substantially lower rates than whites and other minorities, we inquired into what had been done to assure that the two tests we analyzed and other applicant appraisal procedures are job related and valid, that is, that performance on the selection instrument provides information about the probability of success on the job.

We found that a substantial amount of evidence has been gathered to support the validity of the PACE written test. However, because of PACE's impact on blacks and its use for over 100 occupations, OPM should collect additional evidence of its validity and job relatedness or develop substantially equally valid alternative procedures which eliminate adverse impact. Little scientific evidence was available to support the validity of JFA and the unassembled examinations, but OPM is currently revising these procedures.

WHAT IS VALIDATION?

Validation is the process of determining what types of inferences can be drawn from test scores or other forms of

---

1/See GAO report, "Problems with Federal Equal Employment Opportunity Guidelines on Employee Selection Procedures Need To Be Resolved" (FPCD-77-54, Feb. 2, 1978).

personnel assessment. In employment, the question is whether or not the test scores or other assessments provide information about probable future job performance. The Uniform Guidelines recognize that validation is highly technical and complex and that the concept is constantly changing as a result of advances in the field of industrial psychology.

Professional standards for the production and use of tests have developed over the years and have been jointly published by the American Psychological Association (APA), the American Educational Research Association, and the National Council on Measurement in Education. The most recently published professional standards, known as the "APA Standards," were published in 1974.

The Uniform Guidelines provisions relating to validation of selection procedures are intended to be consistent with generally accepted professional standards for evaluating standardized tests and other selection procedures (section 5C). Professional practice and the Uniform Guidelines recognize three types of acceptable validation studies:

- Criterion-related validity. Criterion-related validity is demonstrated by empirical data showing that the selection procedure is predictive of, or significantly correlated with important elements of, work behavior.
- Content validity. Content validity is demonstrated by showing that the content of a selection procedure is representative of important aspects of performance on the job, such as a typing test for a typist job.
- Construct validity. Construct validity is demonstrated by showing that the selection procedures measure the degree to which candidates have identifiable characteristics--constructs--which have been determined to be important for successful job performance. Psychologists use the term "construct" to refer to a general trait that is not observable but is constructed from a network of scientific laws and observations about how people behave. Examples of psychological constructs include "clerical aptitude," "mechanical ability," and "inductive reasoning."

When a validity study is called for by the Uniform Guidelines--i.e., when there is evidence of adverse impact as defined by the Guidelines--the Guidelines prescribe specific minimum technical standards which must be met in conducting the study. In addition to meeting the technical

standards, the study must include an investigation of suitable alternative selection procedures and suitable alternative methods of using the procedure which has the least adverse impact.

#### HAS PACE BEEN PROPERLY VALIDATED?

Several years and over \$2 million have been spent developing PACE and researching its validity. The research staff at OPM believe that PACE is a valid predictor of job performance for the occupations for which it is used. They stated that the validity research meets professional standards, practices, and principles of APA, the Division of Industrial-Organizational Psychology of APA, and professional literature in textbooks and other publications, and that the test conforms to the civil service regulations in effect at the time validity data was collected. They also believe that the test substantially meets the Guidelines requirements, although the Uniform Guidelines may not apply because adverse impact as defined by the Guidelines--the bottom line concept--has not been demonstrated.

#### PACE validation strategy

Validation is an extremely complex and technical concept. A full description of the validation work on PACE would require a lengthy technical discussion and is beyond the scope of this report. Nevertheless, we believe that some explanation of the validation strategy is needed to understand why debate exists among various groups as to whether PACE has been properly validated for all the occupations for which it is used. OPM has published a series of reports and other documents describing in detail how PACE was developed and the research done to infer its validity. These reports are available from OPM for those individuals who desire a full discussion of the PACE validity strategy. (See app. XI for a compendium of published documentation.)

Professional standards hold that judgments of construct validity are based on a preponderance of evidence from a variety of sources. In view of this, OPM researchers prepared summaries of the evidence developed over the last 40 years on the validity of tests like PACE. One such review summarized 126 criterion-related validity studies involving jobs similar to those for which selections are made through PACE. In almost every study, the construct being investigated was a significant predictor of job performance. A series of four reviews, summarizing 132

studies, describes how these constructs have been measured over the years and provides additional statistical evidence of their validity.

The primary objective of developing PACE was to provide a single job-related test battery, or group of related subtests, for selecting people for a large number of professional and administrative occupations at the GS-5 and GS-7 entry levels. A secondary objective was to design the test battery so that scores on the subtests could be differentially weighted for various occupational series.

Civil service researchers decided that construct validation was the most appropriate for PACE. Content validity was believed to be the least relevant method because individuals selected are expected to progress to successful performance in occupations where they previously had no knowledge or experience. The researchers believed that criterion-related validity could not be accepted as the sole strategy for two reasons. First, they believe that the approach would not be technically feasible for the large number of occupations for which the test would be used. Second, criterion-related studies have limited generalizability. Such studies can show whether test scores significantly correlate with measures of job performance, but individually they say little about the proper use of the test for occupations where criterion-related studies are not feasible but where a job analysis can be used to relate test performance to job performance.

Because PACE is primarily an entry level test where prior job experience is neither expected nor required, the test is designed to measure cognitive abilities, such as the ability to read and comprehend material similar to that for the target job, quantitative reasoning, and abstract reasoning. Demonstrating construct validity involves, first, a job analysis or some other evidence showing that the constructs--traits or characteristics--being tested for actually relate to important work behaviors and, second, showing that the test actually measures these constructs.

The basic design of the research which led to development of PACE was to:

- Analyze occupations to determine what duties are performed at the journeyman level (i.e., full performance level).
- Analyze the duties to determine what abilities are important for performing those duties.

--Select test parts which measure these abilities.

--Develop a system of differentially weighting the test parts according to occupational requirements.

At the time research began, over 120 occupations met the PACE scope of coverage, but job analyses were performed for only 27 occupations. These 27 occupations accounted for about 70 percent of prior years' placements. Classification standards for the 27 occupations were analyzed to determine the duties or major job components performed by incumbents working at the full performance level within each occupational series. These duties were reviewed and refined by subject matter experts.

A tentative listing of the knowledges, skills, abilities, and other characteristics that were judged to be required in these occupations was developed. This list, like the duties list, was based on a review of the classification standards. Through a review of the psychological literature, six abilities were identified as having potential for inclusion in the written test portion of PACE. One ability (long-term memory) was later eliminated because testing literature did not contain any tests suitable for use in a short-term testing session.

OPM used 1,241 subject matter experts (generally supervisors) in the 27 occupational series to rate the duties that are performed for their importance to successful performance in the occupation and for the relative amount of time spent in each duty. The subject matter experts also rated the abilities for their importance for successful job performance.

OPM psychologists, experienced in the use of tests for employee selection, rated each of the five abilities for its importance to performing the duties for each of the 27 occupations. For each occupation, the duty-importance, time-spent, and ability-importance ratings were combined mathematically and used to weight the abilities to be measured by the subtests of the battery. Six weighting patterns emerged for the 27 occupations, 2 of which covered 23 occupations. Scores on the PACE subtests were to be multiplied by the weights and the sum of the products used to rank competitors for an occupation.

As soon as the PACE written test was developed, follow-up research was started using criterion-related validity studies to test out the system of identifying and weighting ability constructs which underlie job performance. If the

criterion-related validity studies demonstrated empirically that abilities do indeed underlie job performance, this would then lend support to the entire system. According to OPM researchers, it would then be unnecessary to perform criterion-related validity studies in each specific occupation covered by the examination. Abilities important to successful job performance in those occupations could then be linked to abilities measured by the test through a job analysis of the other occupations.

Three occupations have been studied which, in 2 years preceding the study, accounted for the largest number of placements of the 118 occupations covered by PACE. These occupations were: Social Insurance Claims Examiner, Customs Inspector, and Internal Revenue Officer.

The basic design of the studies was to:

- Conduct a job analysis--i.e., determine what journeymen do on the job.
- Use the job analysis to develop measures of job performance.
- Determine the statistical relationship between the incumbent's performance on the test and performance on the job.

In each occupation a detailed job analysis was conducted. This included asking journeymen in each occupation to identify the tasks they perform. Claims authorizers identified 528 tasks, internal revenue officers identified 260 tasks, and customs inspectors identified 494 tasks.

Journeymen were then asked to indicate whether or not they performed each task and to indicate the relative amount of time spent on each. This rating was made on a seven point relative-time-spent scale ranging from "very much below average" to "very much above average."

Responses to the task inventory were analyzed to determine the relative amount of time spent in performing each task by all journeymen. This was considered the measure of the relative importance of each task. An additional analysis was performed in the customs inspector and claims authorizer samples to determine if all journeymen in the sample were performing similar tasks.

Results of the task inventory were used to develop job performance measures for each occupation. Four measures of job performance were developed--job knowledge tests, work samples, supervisory ratings, and supervisory rankings.

Subject matter experts developed test items designed to measure the job knowledge required to perform the duties on which the journeymen spent the greatest amount of time.

Work samples were designed to be relevant approximations to the work performed on the job. In the claims examiner study the work sample consisted of a standardized claim which had to be adjudicated. The work sample in the internal revenue office study consisted of five taxpayer delinquent accounts in which the revenue officer had to make various collection decisions. For the customs inspector study, a videotape simulation was developed. Four sequences of customs activities were shown, and upon completion of each sequence the inspectors were required to complete appropriate customs documents, identify mistakes made during the televised sequence, and recommend proper performance.

A supervisory rating form was designed to record a first-level supervisor's rating of the performance of the journeymen. The rating scales were developed to correspond to the duties identified in the task analysis. Each supervisor rated his journeymen on different categories of performance for each of the major duties identified in the task inventory. Scale points describing effective and ineffective performance were developed for each scale on the rating form.

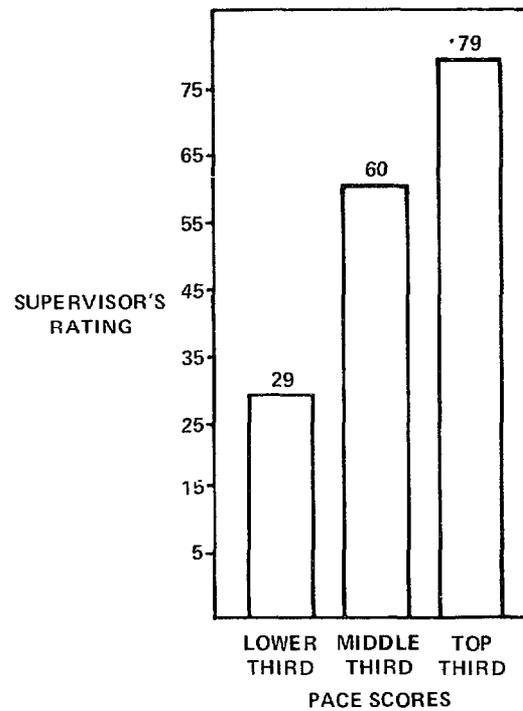
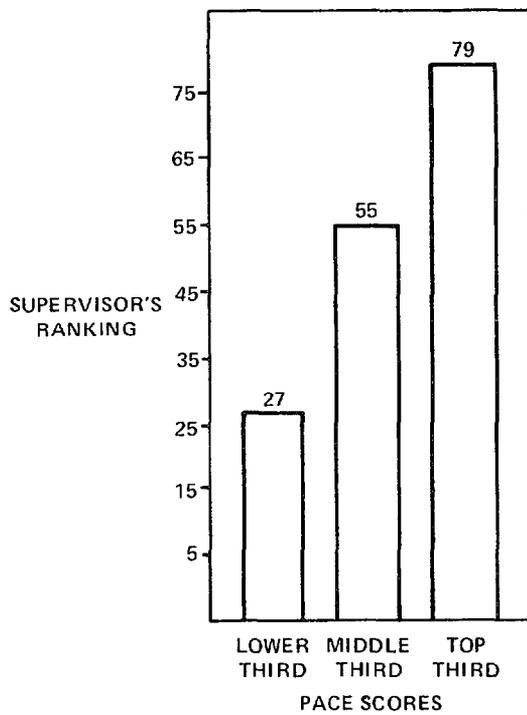
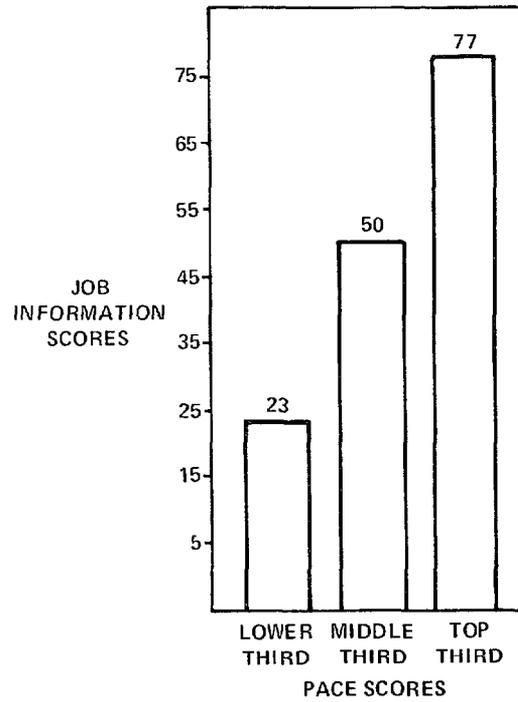
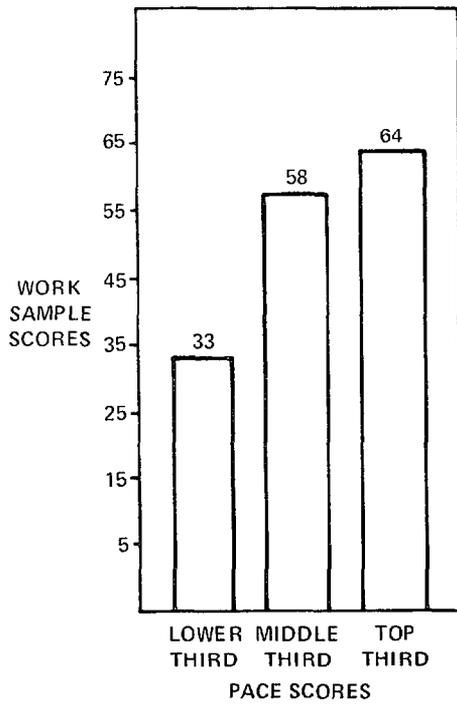
A supervisory ranking form was designed containing the same description of the job duties as the rating form, but without scale points for effective or ineffective performance. Each supervisor had to rank his subordinates with respect to each of the major duties identified for each occupation. This criterion measure was not used in the internal revenue officer study because the officers' union contract forbade such usage.

Two hundred and thirty-one claims authorizers, 305 internal revenue officers, and 190 customs inspectors at various locations throughout the United States participated in the research. They were administered the PACE and the criterion instruments.

### Results of criterion-related studies

The studies showed that the total score on the PACE test was valid as an indicator of job performance on all the measures for claims authorizers and internal revenue officers. For customs inspectors, PACE scores were valid indicators of performance on the job-information test and on the work-sample test but not on the supervisory ratings and rankings. The researchers concluded that the supervisory ratings and rankings were inadequate criteria for measuring performance of customs inspectors because supervisors of the inspectors did not have adequate opportunity to observe subordinate performance.

The charts below--which are illustrative of the relationships found between PACE scores and job performance measures--show the average scores on four performance measures for claims examiners.



As previously indicated, part of OPM's research consisted of administering PACE to groups of research participants and comparing their scores on the test to their scores on the measures of job performance. In addition, the scores on measures of job performance were compared on the basis of how the participants had originally entered the occupation. This comparison, summarized in the table below, showed that research participants who originally entered their occupation as a result of taking PACE or its predecessor, the Federal Service Entrance Examination, scored higher on all measures of job performance than participants who entered the occupations through an alternative selection procedure. The differences were statistically significant for 10 of the 12 comparisons--they were not statistically significant for comparisons of customs inspectors supervisor's ratings and rankings.

Average Percentile Scores For Occupation Entry Method

	<u>Internal revenue officers</u>		<u>Customs inspectors</u>		<u>Claims examiners</u>	
	<u>Test entry</u>	<u>Alternative entry</u>	<u>Test entry</u>	<u>Alternative entry</u>	<u>Test entry</u>	<u>Alternative entry</u>
<u>Job performance measures</u>						
Job information test	56	44	55	39	59	41
Work Sample	58	42	57	37	57	41
Supervisor's rating	54	46	52	46	55	42
Supervisor's ranking	(a)	(a)	50	47	58	42
Composite	59	41	(a)	(a)	(a)	(a)

a/The supervisory ranking form was not used in the research on the internal revenue officer occupation but a composite score was computed. Composite scores were not computed for customs inspectors or claims examiners.

What is the effect of PACE  
on work force output?

OPM researchers attempted to estimate the economic value of PACE in terms of how much its use contributes to improved productivity of the work force in occupations it covers. Using professionally accepted models and equations, they estimated that, under existing conditions, the use of PACE could be expected to enhance Federal productivity. They also estimated the value of PACE if the test were used under varying conditions.

OPM's estimates of the economic value of PACE are based on three relatively simple notions:

- First, the larger the number of applicants relative to the number of vacancies, the more opportunity the employer has to select only those with high ability.
- Second, the more valid the test, the more precisely it identifies ability; consequently, the fewer errors it makes in identifying those with high ability.
- Third, the higher the productivity of the work force selected by some means other than the test, the less likely the test is to make an improvement.

In their analyses, OPM researchers developed a mathematical model and observed the effects of systematically varying (1) the size of the applicant pool relative to the number of vacancies, (2) the validity of the test, and (3) the productivity of those selected by some other means. These analyses produced an upper limit of \$1.2 billion and a lower limit of \$53 million annually. OPM researchers believed, however, that a reasonable estimate of PACE's value was about \$500 million, given existing conditions in terms of the size of the applicant pool, the number of vacancies to be filled, existing data on the validity of the test, and quantitative estimates of work force productivity.

These analyses are based on an assumption that selection is done in terms of rank order on PACE--that is, those with higher scores have a better chance of being selected. The value of the test would be lower if selection were done on some other basis--such as selection at random (e.g., "first in, first hired") from among those who pass the test.

## Opposition to PACE

Despite the extensive research on PACE, and the statements regarding its validity, there continue to be questions raised about its use. For example, in December 1977, the U.S. Commission on Civil Rights stated that, as of August, "the PACE had not been adequately validated \* \* \* to ensure that it was a good predictor of job performance." The Commission recognized the intensive research on PACE, but, in its view, the research available at that time did not provide the necessary evidence of proper validation.

In February 1978, the Civil Rights Task Force of the President's Reorganization Project asserted that PACE "has not been properly validated." On September 8, 1978, the Chairman, CSC--now the Director, OPM--responded to the Task Force by stating that:

"Actually, quite the opposite is correct. The written test used as part of the PACE is the most fully validated and documented test in the Commission's history. We have extensive validation data which we believe satisfies both the relevant professional standards for validation of written tests and published Federal guidance on employee selection procedures. Our staff welcomes opportunities to share that evidence with interested parties."

Further criticism of PACE came in October 1978 from the Department of Justice. In commenting on a plan to form a panel to review all selection procedures for conformance with the new Uniform Guidelines, the Assistant Attorney General, Civil Rights Division, stated that:

"There is however, at least one area in which action ought to be taken, without waiting for the report of such a panel. I make reference to the Professional and Administrative Career Examination (PACE). Our information is that this examination has a severe adverse impact on the applications of blacks and Hispanics. I also understand that the validity studies conducted to date do not satisfy the Uniform Guidelines because criterion-related studies were done for only a few of the more than 100 job classifications covered by the PACE and because those job classifications which were studied do not involve common critical work behaviors with those that were not studied. See Section 14D of the Uniform Guidelines."

As of March 1, 1979, the Director, OPM, had not responded to the Assistant Attorney General's comments regarding the validity of PACE.

In commenting on our report, OPM said:

"\* \* \* this discussion should point out the relative professional and technical resources available in the agencies criticizing PACE and in OPM. While the OPM staff includes a large number of experienced selection psychologists, to our knowledge no such professional and technical expertise is available in the CRC [the Civil Rights Commission], the Civil Rights Task Force of the President's Reorganization Project, or in the Civil Rights Division of Justice. This is a highly relevant item of information to the reader attempting to evaluate technical comments on PACE and its associated studies emanating from these agencies.

"GAO cites only critics from special interest organizations who have often fought against use of standardized tests and have never supported use of such tests. These groups lack the technical expertise to evaluate validity evidence."

The criticisms frequently voiced by those who believe the PACE validation strategy was inadequate are that:

- The construct validity strategy was inappropriate because it is new to the employment field and because an extensive effort is required to support it.
- The written test is used for 118 occupations, but detailed job analyses identifying the work behaviors required for successful job performance were conducted for only 27 occupations.
- Criterion-related studies were performed for 3 of the 12 to 15 occupations where they might have been technically feasible.
- The studies used the concurrent design rather than the predictive design, although predictive studies are preferred from a scientific standpoint.

- Race and ethnic data on the research participants was not collected or analyzed, and civil service researchers did not assure that individuals in the study groups were, to the extent feasible, representative of the relevant labor market.
- Possible test bias or test unfairness for race, sex, or ethnic subsamples was not investigated.
- Suitable alternatives to the test or alternative ways of using the test were not sufficiently investigated.

#### Appropriateness of construct validation

Critics who question the appropriateness of using the construct method of validation point out that the concept is new in the employment field. They cite, for example, the Uniform Guidelines, which state that:

"Construct validity is a more complex strategy than either criterion-related or content validity. Construct validation is a relatively new and developing procedure in the employment field, and there is at present a lack of substantial literature extending the concept to employment practices. The user should be aware that the effort to obtain sufficient empirical support for construct validity is both an extensive and arduous effort involving a series of research studies, which include criterion-related validity studies and which may include content validity studies. Users choosing to justify use of a selection procedure by this strategy should therefore take particular care to assure that the validity study meets the standards set forth [in the Guidelines]." (Section 14D(1))

Construct validity requires the commitment of substantial resources and considerable expertise. However, it also has considerable payoff as a cost-effective selection device when:

- There are a large number of jobs with some common work behaviors.
- The employer needs to select large numbers of employees.

--A test is needed to measure abilities, aptitudes, or traits for which content validity is not appropriate.

For these reasons, construct validation, when used, has been most widely used in mass selection situations, particularly in the Government. This method was used to develop the Army Alpha test of intelligence used during World War I, the Aircrew Classification Battery used during World War II, and the Armed Services Vocational Aptitude Battery which is currently being used. It was also used to develop and validate the General Aptitude Test Battery used for vocational guidance and job referral by the U.S. Employment Service, and it has been the basis for various aptitude test batteries--such as PACE--used for civil service selection.

PACE is the type of ability test that has been extensively investigated over the last 40 years, and for which a substantial amount of criterion-related and construct validity evidence has accumulated. Civil service researchers relied on that accumulated research literature in developing PACE.

We believe that construct validation was an appropriate technique to use in developing and researching PACE. OPM has had the technical resources to develop and validate tests using the construct method of validation, and professional care and judgment was generally exercised in developing and researching the test. However, standards relating to the need for a representative sample of research participants and investigations of test fairness were not followed. These standards are described as "essential" <sup>1/</sup> by the American Psychological Association and the Uniform Guidelines. Also, job analyses called for by the American Psychological Association and the Uniform Guidelines were not performed for all occupations for which the test is used.

In our opinion, if the test had less of an adverse impact on blacks, or was less important in obtaining a professional or administrative job in the civil service, then

---

<sup>1/</sup>According to the American Psychological Association, the "APA Standards" were prepared as a technical guide for those within the profession and were not written as law. The APA Standards state that it is undesirable to treat the standards as unduly rigid, and that, "Standards are statements of ideals or goals, some having priority over others. \* \* \* An evaluation of competence depends on the degree to which the intent of this document has been satisfied by the test developer or user."

the validation strategy and evidence supporting the test would be very persuasive. We believe, however, that in view of the impact the test has on black applicants, and its importance in obtaining a civil service job, each of the professional standards and Uniform Guidelines requirements must be scrupulously followed.

Limited scope of job analysis  
and criterion-related studies

Intensive and detailed job analyses were performed on 27 occupations during the initial research and development work on PACE. Job analyses of the same intensity were not performed on the remaining 91 of the 118 occupations for which PACE is used.

In commenting on our report, OPM said that the criticism concerning the lack of job analyses on the remaining 91 occupations was unfounded and that job analyses had been performed on all 118 occupations.

In considering OPM's comment, we noted that initially over 120 occupations met the PACE scope of coverage requirements. Because the occupational categories were developed on a sample of 27 occupations, OPM decided that it was necessary to determine into which categories the remaining occupations should be placed. Therefore, during 1974, as an interim measure, an occupational specialist familiar with PACE, and with the qualifications and classification standards of occupations in its scope of coverage criteria, evaluated each of the remaining occupations in the context of the five abilities tested by PACE. The occupational specialist then placed each occupation into an existing occupational category pending the outcome of a planned comprehensive analysis. This interim process resulted in several occupations being deleted from PACE coverage because they did not meet the scope of coverage criteria, and other occupations being included because they did meet the criteria. From time to time, other modifications to PACE coverage have been made as a result of standards work.

The planned comprehensive analysis of the remaining occupations was to have involved subject matter experts rating the importance of duties and time spent on each duty for their own occupations, and also rating the importance of the abilities for performance of duties. Weighting patterns were then to have been established as was done for the 27 occupations. While some additional work was done with respect to performing job analyses on the

remaining occupations, available documentation did not demonstrate that the comprehensive analysis initially envisioned was completed. Also, documentation was not available to show that the jobs studied during the research and development of PACE involved critical work behaviors common to those that were not studied. In this regard, the Uniform Guidelines state:

"There should be a job analysis. This job analysis should show the work behavior(s) required for successful performance of the job, or the groups of jobs being studied, the critical or important work behavior(s) in the job or group of jobs being studied, and an identification of the construct(s) believed to underlie successful performance of these critical or important work behaviors in the job or jobs in question. Each construct should be named and defined, so as to distinguish it from other constructs. If a group of jobs is being studied the jobs should have in common one or more critical or important work behaviors at a comparable level of complexity." (Section 14D(2))

The Guidelines also describe the type of documentation which is to be maintained for each job covered by the selection procedure:

"A description of the method used to analyze the job should be provided (essential). A complete description of the work behavior(s) and, to the extent appropriate, work outcomes and measures of their criticality and/or importance should be provided (essential). The report should also describe the basis on which the behavior(s) or outcomes were determined to be important, such as their level of difficulty, their frequency of performance, the consequences of error or other appropriate factors (essential). Where jobs are grouped or compared for the purposes of generalizing validity evidence, the work behavior(s) and work product(s) for each of the jobs should be described, and conclusions concerning the similarity of the jobs in terms of observable work behaviors or work products should be made (essential)." (Section 15D(4))

While some analysis was performed by OPM to determine which occupations met the PACE scope of coverage, it appears that additional work is needed to satisfy the Guidelines

requirements with respect to performing job analyses and assuring that the jobs covered involve common critical work behaviors. OPM told us that it now has additional studies underway for creating a basis for generalizing across jobs based on common duties.

The professional standards and the Uniform Guidelines do not specify how many criterion-related validity studies are needed to demonstrate empirically that abilities tested for do indeed underlie job performance. The Guidelines, however, call for a series of criterion-related validity studies:

"If construct validity is to be generalized to other jobs or groups of jobs not in the group studied, the Federal enforcement agencies will expect at a minimum additional empirical research evidence meeting the standards of subparagraphs section 14B(2) [analysis of the job] and (3) [criterion measures] above for the additional jobs or groups of jobs." (Section 14D(4))

OPM researchers told us that criterion-related studies may be technically feasible in 12 to 15 occupations. In view of the impact which PACE has on job opportunities of black applicants, we believe that some additional criterion-related studies should be performed to support the construct validity of the test.

#### Use of predictive versus concurrent studies

There are two basic designs for performing criterion-related studies--predictive and concurrent. Predictive studies involve:

- Administering the selection procedure to applicants.
- Hiring individuals without regard to scores on the selection procedure.
- Obtaining criterion measures at a later date.
- Determining the degree of relationship between the selection instrument and the criterion--i.e., the measure of job performance.

Concurrent studies, on the other hand, involve:

- Administering the selection instrument to a group of current employees.
- Obtaining criterion measures on that group.
- Determining the degree of relationship between the selection instrument and the criterion.

The validation principles of the American Psychological Association, Division of Industrial-Organizational Psychology, state that a predictive study is preferred from the standpoint of scientific merit because its use properly begins with job candidates, not job incumbents. The concurrent method cannot be expected to answer questions of prediction. It can only answer questions about relationships of a given characteristic of preselected employees at a given time. In theory predictive studies permit the analysis of the job performance of persons who do not score well on written tests.

The distinction between predictive and concurrent studies is important because of the possibility that the relationship found between test scores and performance for current employees may not be the same relationship which would be found for a group at random. In practice, however, this rarely happens, and when the two groups differ there is usually a shorter range of ability in the current employee group. The reason for this is that by the time a concurrent validation study is conducted, those with lesser ability may have been fired or quit and those with more ability may have been promoted to other jobs. Consequently, a test which is a valid measure of these abilities will be found to have a correlation with performance among present job incumbents which is smaller than the correlation it would have for the group selected at random.

OPM researchers found no restriction in the range of test scores of the groups who participated in the research on PACE. The variances of their test scores were not significantly different from the test scores of PACE applicants.

OPM researchers told us that, in their opinion, predictive validity studies were not technically feasible because civil service laws require that applicants be ranked according to their fitness for the job, and applicants cannot be hired without regard to their scores on the selection procedure. This opinion, of course, presupposes that the selection procedure being validated will fairly rank applicants according to their fitness for the job, that is, the procedure is already valid and job related.

While a predictive validity study may not be technically possible in a system requiring merit selection because all applicants cannot be hired without regard to scores, it may be possible to carry out a research study where all applicants selected by some other merit procedure are administered the test and then followed up at a later date. Some of the other merit procedures currently used to fill PACE and other jobs include Cooperative Education Programs, the Presidential Management Intern Program, and the Vietnam-Era Veterans Readjustment Program. (See ch. 4.) In our opinion, a research study in which predictor information is obtained prior to placement of employees on a job and criterion information obtained later would more convincingly answer the most common employment question: Does the test have predictive value with respect to later job behavior?

In commenting on this report, OPM said that, while the type of study suggested above has certain drawbacks, it is planning such a study with the Social Security Administration. OPM said that, to the extent feasible, it will assure that the study participants are representative of the relevant labor market and it will investigate test fairness.

Collection of race and ethnic data  
and assessment of test fairness

The American Psychological Association has stated that the following standard is essential in performing criterion-related validity studies.

"The sample should be described in terms of those variables known as thought to affect validity, such as age, sex, socio-economic status, ethnic origin, residential region, level of education, or other demographic or psychological characteristics."

Similarly, the Uniform Guidelines state that:

"Whether the study is predictive or concurrent, the sample subjects should insofar as feasible be representative of the candidates normally available in the relevant labor market for the job or group of jobs in question, and should insofar as feasible include the races, sexes, and ethnic groups normally available in the relevant job market." (Section 14B(4))

We noted that in performing the criterion-related validity studies the OPM researchers did not endeavor to assure that the research participants were representative of the relevant labor market. Biographical information was gathered concerning the participants' age, sex, method of selection into the occupation, work experience, education, training, and other personal data, but information on race and ethnicity was not obtained. The significance of data on race and ethnicity is that it permits investigating questions of test fairness.

The Uniform Guidelines recognize that the concept of test fairness or unfairness is a developing concept. Nevertheless, the Guidelines state that when a selection procedure results in an adverse impact on a race, sex, or ethnic group, and that group is a significant factor in the relevant labor market the test user generally should investigate the possible existence of unfairness for that group if it is technically feasible to do so. The greater the severity of the adverse impact on a group, the greater the need to investigate the possible existence of unfairness. (Section 14B(8)(b))

The Guidelines define unfairness as follows:

"When members of one race, sex, or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences in scores are not reflected in differences in a measure of job performance, use of the selection procedure may unfairly deny opportunities to members of the group that obtains the lower scores." (Section 14B(8)(a))

The Guidelines state that test users conducting a study of test fairness should review the American Psychological Association standards regarding investigations of possible bias. The American Psychological Association standard regarding test bias states that:

"A test user should investigate the possibility of bias in tests or in test items. Wherever possible, there should be an investigation of possible differences in criterion-related validity for ethnic, sex, or other subsamples that can be identified when the test is given. The manual or research report should give the results for each subsample separately or report that no differences were found."

The possibility of test bias or unfairness was not investigated as part of the research on the validity of PACE. The researchers told us that the notion of so-called differential validity or unfairness has been thoroughly discredited scientifically. A substantial amount of research published since about 1972 has fairly consistently shown that tests demonstrated to be valid are equally valid for both majority and minority groups; that is, groups which perform less well on the tests of performance tend to do less well on the measures which the tests are designed to predict.

The question of whether a valid test can be biased or unfair is often raised. The answer to that question depends upon which of the competing definitions of bias is used, each of which incorporates different social values.

- One definition holds that bias exists when test scores prevent deserving groups from being adequately represented among selected applicants. This is a quota-based definition and does not allow for selecting on the basis of job-related traits when such selection would lead to disproportionate representation among applicant groups.
- Another definition holds that bias exists when those with equal chances of success on the job have unequal chances of being selected for the job. The question of whether tests are equally valid predictors for the majority and minority group is relevant to this definition. This definition maximizes productivity and is based on competition or merit and equal opportunity for the individual.
- A third definition holds that bias exists when the number of applicants from a certain group selected by the test is smaller than the number from that group who could perform satisfactorily. This definition differs from the second definition in that it is focused on equality of opportunity for groups rather than individuals, and on actual rather than predicted performance.

The published literature on test validity indicates that most tests are either fair to minority groups or slightly biased in their favor by the second definition, which, according to OPM, is the only concept of fairness consistent with merit system principles. The literature also indicates that, by the third definition, tests are

slightly biased against minority groups, and if one subscribes to the first definition, tests have always been biased against minorities.

#### Investigation of alternatives to PACE

The Uniform Guidelines state, in effect, that Federal EEO law has added a new requirement to the traditional process of validation. The Guidelines state that, whenever a validation study is called for, the study should include an investigation of suitable alternative selection procedures and suitable alternative methods of using the selection procedure which have as little adverse impact as possible. The employer cannot concentrate solely on establishing the validity of the instrument or procedure which it has used in the past. The alternative selection procedure--or method of use--should be used when it has less adverse impact and when the evidence shows that it is substantially equally valid for the same job in similar circumstances.

OPM researchers said that alternatives to PACE, or alternative ways of using PACE, were explored during their research efforts. For example, they investigated various ways of using college majors or grade-point average, and specifically tailored written tests as alternatives for selecting individuals for PACE occupations. Their research indicated that these alternatives were not feasible.

Because of the impact PACE has on minorities, the OPM Regional Office in Dallas, Texas, participated with the Department of Health, Education, and Welfare in a study to find suitable alternative procedures for selecting entry level social security claims examiners. The study began in July 1976 and, while it was not specifically a part of the validation research on the PACE written test, it was reviewed by the OPM researchers.

A task force was established in the Dallas Region with the stated objectives of developing an alternative selection procedure that was (1) based on sound merit principles, (2) designed to facilitate EEO affirmative action goals, and (3) practical to administer. As of December 1978, the task force had not identified an alternative procedure which met these criteria for selecting entry level claims examiners.

Other alternative methods being used to bring individuals into PACE occupations are discussed in chapter 4. However, those procedures were not developed or investigated

as part of the validity research on PACE, and consequently, data is not available to indicate whether any of the alternative selection methods are of substantially equal validity.

Some work has been done to explore alternatives to PACE which have less of an adverse impact on black applicants. However, we believe that more can and should be done to investigate these and other alternatives which may be equally valid.

In commenting on this report, OPM said:

"In considering alternative procedures, it is important to recognize that the most common feature of PACE occupations \* \* \* is their information burden. Many jobs not labeled as "Analysis," "Investigation," "Examining," or "Inspection" positions relate either to a specific profession (General Schedule 110, 150-193, 950) or to topic areas for which the label "specialist" implies mastery of a body of knowledge. There exists a complex written body of laws and regulations which affects each PACE job. The cognitive abilities measured by PACE are, therefore, clearly related to success in these occupations and it would be illogical to assume that ways may be found to bypass consideration of these abilities and still to ensure competence in these occupations.

"This fact--that PACE jobs are information-handling jobs--is a bedrock feature of the selection problems relating to PACE. Although there are additional job-related skills not measured by the PACE due to feasibility problems (character variables, personality variables and skills mentioned such as long-term memory), the necessity for paying close attention to information skills (the "constructs" measured by Test 500 clearly relate to the intake, evaluation and processing of information) cannot be ignored. There is no known cost-effective means of measuring them other than with PACE or a similar instrument. We strongly recommend that the report recognize these realities." (OPM's underscoring)

The EEOC, in commenting on this report, said:

"The most important single failing of CSC's approach to selection procedures is the failure to adequately seek alternative methods of selecting and an equally significant failure to encourage individual agencies to develop and identify alternative selection procedures which will reduce or eliminate adverse impact while continuing to meet legitimate merit system goals."

HAS THE JFA TEST BEEN  
ADEQUATELY VALIDATED?

The written test used as part of the JFA selection procedure was developed in 1966 and, according to OPM, it was developed in accordance with the then applicable professional standards and practices for construct validity.

The written test is used for entrance into 28 technical support and clerical occupations at the GS-4 level, and OPM researchers believe that it is a valid predictor of job performance. However, they advised us that the documentation available to support the validity of the test probably does not meet current regulatory requirements.

The Uniform Guidelines state that criterion-related validity studies described in the professional literature will be considered acceptable under certain conditions. To comply with the Guidelines to the extent possible, OPM recently reviewed the published and unpublished criterion-related studies relative to the eight different item types used in the JFA written test to measure abilities for jobs similar to the occupations for which the test is used. According to OPM, validity data was collected on 13 occupations. No attempt was made to locate data for the other 15 occupations because the data on the 13 occupations supported the use of the test for 28 occupations since they all have similar ability requirements.

OPM recently started a project to develop and document the validity of a new written test to replace the test currently used for JFA occupations. This project will include a job analysis to determine which work behaviors are required for successful job performance, and a list of measurable knowledges, skills, and abilities will be developed and linked to work behaviors. The results of the job analysis will guide the research team in determining the appropriate validity strategy to use. If the job analysis shows that the important work behaviors and knowledges, skills,

and abilities can be captured by a selection procedure which will closely approximate an observable work behavior, then a content validity strategy will be used to validate the test. However, if the job analysis indicates that a test should be devised to measure work behaviors and knowledges, skills, and abilities which are more abstract, then a construct validity strategy will be followed.

OPM plans to finalize the new JFA selection procedures by October 1981. The validations work is planned to be completed and documented by October 1982.

HAVE UNASSEMBLED EXAMINATIONS  
BEEN ADEQUATELY VALIDATED?

To satisfy merit system principles, all civil service selection procedures are required to be objective, reliable, valid, and job related. If the selection procedures for a given job have an adverse impact on minorities, then the provisions of the Uniform Guidelines come into play. These requirements are applicable to unassembled examinations--evaluations of relevant education, training, and experience--as well as to written tests.

Our review did not include a specific inquiry into what research had been performed to assure that the selection procedures used for the Accountant-Auditor and Social Worker occupations are valid and job related. However, we reviewed the actions taken to resolve the problems concerning unassembled examinations discussed in our report, "Improvements Needed in Examining and Selecting Applicants for Federal Employment" (B-179810, July 22, 1974). In that report we found that the reliability of CSC's unassembled examining procedures needed improvements so that applicants' scores could be more consistently determined.

OPM reported several actions taken to improve the reliability of the unassembled examining process. These included:

- Improving the clarity and usefulness of rating guidelines.
- Developing a training program for rating examiners to increase their proficiency.
- Developing a quality assurance program to provide continuing and statistically valid review of the ratings.

We were told that the type of documentation needed to demonstrate the validity of unassembled examinations in accordance with professional standards was never developed. The job relatedness of the qualification standards developed for an occupation, and the rating schedules used to rank applicants, were generally justified as being job related on the basis of reasonableness rather than on the basis of the type of validity evidence described in the Uniform Guidelines.

In June 1977 a new "Exam Preparation Manual" was published for use by personnel staffing specialists. The Manual describes a procedure for conducting a selection-oriented job analysis and for using the results of the job analysis to develop unassembled examinations. The procedure, known as the Behavioral Consistency Method, is designed to improve and demonstrate job relatedness for unassembled examinations.

The first principle of the new method is that candidates should be evaluated only on those knowledges, skills, abilities, and other characteristics that show the largest difference between superior and minimally acceptable employees. This is the same as saying that candidates should be evaluated only on the "most valid" knowledges, skills, and abilities. The importance is that, for any job, there are large numbers of knowledges, skills, abilities, and other characteristics that are "job related" and show some difference between the best and the poorest performers. But many of these show only small differences and are not useful for rating, and, in the final analysis, only a limited number can be measured. The best strategy to maximize the validity and usefulness of the rating method is to include only 5 to 8 knowledges, skills, abilities, and other characteristics that best separate the superior employees from the minimally acceptable ones.

Once the most valid set of knowledges, skills, and abilities is identified, the next question is: What is the best way to measure applicants? Traditional rating methods attempt to assess an applicant's knowledges, skills, and abilities by evaluating and crediting education and experience. The assumption has been that education and experience lead to the development of job-related knowledges, skills, and abilities. For some applicants, this assumption is probably accurate, but for others education and experience may merely be passive exposures. For example, of 10 candidates who each have 5 years experience as personnel administrators, some may have performed at a very high level and been responsible for many improvements

and changes in personnel procedures, while others may have done only enough during the 5 years to avoid getting fired. But if the amount of experience is all that is credited, all 10 applicants get the same number of points, since all have 5 years of experience.

The next question is how to obtain the information needed about the applicant's achievements. According to OPM there is really only one usable source: the applicant, who has information in the necessary detail about all past accomplishments and achievements. Others may have some information about some of the applicant's achievements, but rarely will they have complete information. Research evidence indicates that people are generally honest in giving written reports of specific, concrete facts about themselves and their achievements. Therefore, the Behavioral Consistency Method requires the applicant to provide specific achievements in narrative form for each job-related knowledge, skill, ability, and other characteristic. For each achievement, the applicant is asked to provide the name, address, and phone number of someone who can verify the achievement. These claims are verified by the hiring agency prior to the final selection decisions.

Once the achievements are collected, how should they be scored to produce a rank order? In most cases, applications must be evaluated by staffing specialists, and the staffing specialists need some sort of guide or scoring aid if their judgments are to reflect achievements as derived from actual applicant responses. The credit value of these benchmarks is determined by subject matters experts.

The Behavioral Consistency Method is based on content validity. The behaviors sampled in the achievements are content valid because they sample the kinds of achievements required in performance on the job. OPM believes that when this methodology is used to develop selection procedures, and the procedures are used properly, agencies will observe a marked improvement in the quality of applicants referred to them from OPM job registers.

Examination packages for five occupations were developed in accordance with the new Manual. These are being tested in five OPM regional offices during fiscal year 1979. OPM plans to have five additional examinations packages completed by March 1979 that will also be tested in various regions.

## CONCLUSIONS

The Government, as an employer, is obligated to operate under the merit system principles which require that tests and other applicant appraisal procedures be valid and job related.

According to OPM, PACE is the most fully validated and documented test in civil service history. Civil service research indicates that persons who score high on the PACE tend to perform better on the job. Although the research indicates the validity and job relatedness of PACE, we believe that OPM must deal with the concerns expressed about the research strategy. In view of the importance of the test and the impact it has on black applicants as shown by our analysis, we believe that the research strategy must scrupulously adhere to both the professional standards and the Uniform Guidelines on Employee Selection Procedures.

The documentation supporting the validity of the written test used as part of the JFA selection procedures does not appear to be sufficient to meet current requirements. Accordingly, OPM is in the process of developing and documenting the validity of a new written test to replace the test used at the time of our review. If the new test is developed using the construct validity approach, we believe a special effort must be made to deal with the criticisms leveled at the PACE validation strategy.

Unassembled examinations should meet the same standards as written examinations of being objective, reliable, valid, and job related. Actions were taken as a result of our 1974 report 1/ to improve the reliability of the unassembled examining process, but the documentation described by the Uniform Guidelines to demonstrate the validity of the unassembled examining process has never been developed. We believe that the procedure described in OPM's Exam Preparation Manual and currently being tested represents an improvement over traditional methods for evaluating education, training, and experience.

OPM advised us that in October 1978 it had begun planning for a comprehensive review of those Federal policies and activities which may be affected by the Uniform Guidelines. Areas of review will include job qualifications standards, examinations, performance appraisals, and promotion policies. Such a review should:

---

1/"Improvements Needed in Examining and Selecting Applicants for Federal Employment" (B-179810, July 22, 1974).

- Determine whether there are conflicts between Federal selection programs and the Guidelines.
- Establish a plan of action for conformance, including recommendations for resources to eliminate any identified problems.
- Recommend continuing current or adopting new policies, methods, and techniques which would help to assure that future policies and practices of OPM will be regarded as being in compliance with the Guidelines by EEOC and the Department of Justice.

#### RECOMMENDATIONS

We endorse the planned comprehensive review of all the Federal policies and activities which may be affected by the Uniform Guidelines. However, we believe attention to PACE cannot await the outcome of the comprehensive review. Accordingly, we recommend that the Director, OPM, reevaluate the PACE validation strategy to assure that it is in conformance with the professional standards and the Uniform Guidelines requirements, including the requirement that substantially equally valid alternatives with less adverse impact be investigated. We also recommend that OPM assure that the comprehensive job analysis initially planned for the occupations not included in the original research and development of PACE be completed and documented, and assure that jobs covered by PACE involve common critical work behaviors.

If further validation studies are necessary, we recommend that the Director, OPM:

- Reconsider the feasibility of performing criterion-related studies using the predictive design, or of performing research where all applicants selected for a PACE occupation by some other merit procedure are administered the test and followed up at a later date.
- Require that race, sex, and ethnic data be collected on research participants, and that every effort be made to assure that research participants are representative of the applicant population, and that the fairness or unfairness of the test be investigated.

## CHAPTER 4

### ARE ALTERNATIVE JOB

#### ENTRY METHODS AVAILABLE?

The primary entry route for new persons into the competitive civil service is through appointment from civil service registers developed from assembled and unassembled examinations and maintained by OPM. However, appointments from civil service registers represent only a portion of the civil service vacancies that are filled annually. Agencies can use other external and internal staffing programs which they control to fill vacancies. Decisions about whether to use CSC registers or an alternative selection procedure are usually based on the agency's overall staffing objective, budget conditions, and affirmative action goals. Agencies have been encouraged to make use of alternative selection procedures when problems are encountered in obtaining a sufficient number of minority or female job candidates from civil service registers.

#### PROBLEMS IN OBTAINING MINORITIES FROM REGISTERS

Personnel officers in several agencies told us that registers developed from written tests do not provide enough minority job candidates to carry out meaningful affirmative action programs. They said that minority candidates were seldom certified from civil service registers that required applicants to pass written tests. For example:

--A Department of Health, Education, and Welfare regional official said, "Minorities, especially blacks, are hard to come by on the CSC assembled examination registers. Judging by the number of blacks certified to HEW, blacks must be failing these exams at an all-time record high."

--A Department of the Air Force activity scheduled interviews with 159 individuals certified from the PACE register, not one of whom was a minority. In a report the Air Force noted, "\* \* \* CSC certificates are the least productive way of selecting minority candidates \* \* \*."

--The Federal Aviation Administration, in setting its equal employment opportunity goals for fiscal year 1976, noted that a major part of its hiring activity

was normally filled from CSC registers. However, such registers contained almost no minorities.

Other agencies indicating few minority referrals from PACE registers included the Department of the Army and the Social Security Administration. In 1974, before PACE was used, the Army appointed about 12 percent minorities to PACE-type positions. In 1976, after PACE was introduced, the number of minority appointments from registers dropped to 1 percent. The hiring rate of minorities in the Social Security Administration dropped from 16 percent before the PACE test was used to 3 percent after the PACE examination was introduced. The Air Force reported substantially fewer minorities being brought into entry level positions since the inception of PACE.

#### ALTERNATIVE STAFFING PROCEDURES

The work force within any Federal occupation is the product of several external and internal selection procedures. OPM has information to show how many individuals are selected for specific occupations from civil service registers, but since alternative procedures are managed by individual agencies, it does not have information to show how many individuals are selected for specific occupations using each of these alternative procedures.

The following chart, while it may not be all inclusive, illustrates some of the various ways an individual may gain entry into an occupation covered by the PACE written test.

INTERNAL PROMOTION FROM A  
RELATED TECHNICAL OCCUPATION

PROMOTION FROM AN AGENCY  
UPWARD MOBILITY PROGRAM

LATERAL REASSIGNMENT FROM  
A RELATED OCCUPATION

REINSTATEMENT OF FORMER  
FEDERAL EMPLOYEES

VIETNAM-ERA VETERANS  
READJUSTMENT PROGRAM APPOINTMENTS

PRESIDENTIAL MANAGEMENT  
INTERN PROGRAM APPOINTMENTS

COOPERATIVE EDUCATION  
PROGRAM APPOINTMENTS

OBTAIN A CERTIFIABLE SCORE ON  
PACE AND BE SELECTED FROM THE  
REGISTER

PACE OCCUPATION; GS-5 OR GS-7 ENTRY LEVEL

Data collected as part of the validation studies on the three PACE occupations--Customs Inspector, Internal Revenue Office, and Social Insurance Claims Examiner--indicated that about half of those individuals who participated in the research entered the occupations from taking PACE or its predecessor, the Federal Service Entrance Examination. The other half entered the occupation from some alternative selection procedure, mostly by promotion from other jobs within the agency. OPM has estimated that about two-thirds of all career or career-equivalent positions are filled annually by individuals who already have a Federal job. The remaining one-third are filled by individuals who have never held a Federal job.

Upward mobility is the major formalized program designed to develop current employees' skills and abilities so that they can meet the qualification standards for target positions in career fields offering advancement potential. Through upward mobility programs, agencies restructure jobs and establish trainee and "bridge" positions to help secretarial, clerical, and technical employees move into professional occupations. Federal agencies are required to develop upward mobility program information as a part

of their EEO affirmative action plans. There were approximately 74,000 upward mobility program participants during fiscal year 1977.

### Noncompetitive job entry programs

For individuals who have never held a Federal job, there are few opportunities for permanent Federal employment except through appointment from a civil service register. However, limited noncompetitive appointments are available to some individuals which can later be converted to competitive civil service status, generally without taking the applicable civil service examination. Among these programs are the Vietnam-Era Veterans Readjustment Program, the Presidential Management Intern Program, and the Cooperative Education Program.

#### Vietnam-Era Veterans Readjustment Program

This program was established by Executive Order 11521 on March 26, 1970. The provisions of the Order were incorporated into the Vietnam-Era Veterans' Readjustment Assistance Act of 1974 (Public Law 93-508). The program is designed to promote the maximum employment and job advancement opportunities within the Federal Government for qualified disabled veterans and veterans of the Vietnam-era. Under this program agencies are authorized to make noncompetitive excepted appointments to a job in the competitive service in grades 1 to 7--general schedule and wage grade--which combine work with self-development training or education. The program includes a combination of planned on-the-job training, classroom training, basic or remedial education, high school or high school equivalency, or education beyond high school. The noncompetitive appointments are converted to career or career-conditional as soon as 2 years of service are completed, provided the veteran's performance on the job is satisfactory and he or she has participated in the agreed-to training or educational programs. A career-conditional appointment leads to a career appointment after 3 years of continuous service.

This program was originally scheduled to end on June 30, 1978, but Public Law 95-520 (Oct. 26, 1978) amended the 1974 Act, extending the authority for making appointments to September 30, 1981. The amendment also removed the 1-year time limit for obtaining an appointment, the 14-year education restriction for compensably disabled veterans and veterans discharged because of service-connected disabilities, and changed the category of discharge required

for an appointment from "under honorable conditions" to "other than a dishonorable discharge."

OPM estimated that during fiscal year 1978 there were about 14,000 hires under this program. However, data was not available on the number hired by occupational category, or on the race, sex, or ethnicity of those selected.

#### Presidential Management Intern Program

This program was established by Executive Order 12008 on August 25, 1977. The program provides a special means of entry into the Federal service for persons who have received, or expect to receive during the academic year, a graduate degree with a concentration in public management. Eligible students must be nominated by their school and successfully complete a regional screening process. Each year 250 interns are selected for the 2-year internships program in Federal agencies. Initial appointments are made at the GS-9 level. Upon satisfactory completion of the 2-year internship, interns are eligible for noncompetitive conversion to career or career-conditional appointments within the same agency.

One goal of the program was to increase the participation of women and minorities in determining public policy. University nominating officials are expected to identify highly qualified women and minorities who show potential for management development. In the first year of the program about 20 percent of the interns were minority group members and 46 percent were women, 44 percent of whom were minority women. Interns are often appointed to occupations for which the PACE written test is used. Interns are not required to take the PACE examination.

#### Cooperative Education Program

According to OPM, some type of a cooperative education program has been in existence since 1946, authorized under various Executive orders. The October 26, 1977, Executive Order 12015 currently authorizes OPM to establish a career-related work study program permitting students completing the program to be noncompetitively converted to career or career-conditional positions in the competitive civil service. OPM has established cooperative education programs for graduate students, baccalaureate students, and students enrolled in 2-year associate degree programs.

At the present time this provision applies only to civilian agencies. The Conference Report (No. 95-1764) accompanying the Department of Defense Appropriations Act, 1979 (Public Law 95-457), requires students participating in cooperative education programs managed by the Department of Defense to pass the appropriate examinations with a score of 70 prior to being given career appointments.

During fiscal year 1978 there was a total of 12,498 students participating in cooperative education programs throughout the Government, an increase of 2,045 over 1977. During fiscal year 1977, 1,538 students completed the program, 916 of whom were given career service appointments. Although specific data is not available, many of the participating students were in occupations for which the PACE written test is used.

#### USE OF ALTERNATIVE METHODS TO ACHIEVE EEO OBJECTIVES

The various alternative staffing programs were established to achieve a variety of goals. Because of their flexibility and the fact that individuals can be hired without taking a competitive entrance examination, it is generally believed that alternative programs are useful tools in hiring more women and minority group members. Information is not available to show how useful the programs have been to increase the number of minority group members in the Federal work force or in occupations for which PACE is used. Agencies generally have not collected and reported data by occupational series and grade level on the race, sex, or ethnicity of individuals selected using alternative procedures. We believe that such information must now be maintained to comply with section 4 of the Uniform Guidelines on Employee Selection Procedures.

The use of alternative selection programs for filling entry level positions is controversial. Some argue that these programs should only be used when there is an insufficient number of qualified job candidates on civil service registers. They argue that not only are these programs costly and time consuming to operate, but extensive use of these programs would undermine the merit system principles of (1) having competitive examinations open to everyone, (2) giving everyone an equal opportunity to be examined for positions for which he or she has the minimum qualifications, and (3) selecting from among the best qualified. Some have also argued that the alternative selection procedures, because of the lack of objective measures, make it easier to discriminate on the basis

of race, sex, religion, national origin, or other factors not directly related to the requirements of the job or jobs involved.

Neither OPM nor the agencies know whether individual alternative selection procedures result in hiring individuals who perform better or worse on the job than individuals selected from civil service registers. If, for example, the PACE written test is more valid and job related than the alternative selection procedures, an individual selected for a job from the PACE register could be expected to perform better than an individual selected for the same job using alternative procedures. Conversely, if individuals selected by alternative procedures perform better on the job than individuals selected from the PACE register, it would appear that the alternative procedure would be more valid and job related than the PACE test.

The alternative procedures are generally thought of as having less of an adverse impact than the PACE written test. However, statistics are not available to show the impact that alternative selection procedures have on minorities. The Uniform Guidelines (sections 3B and 15B (9), C(6), and D(8)) require that, as part of a validity study, suitable alternative selection procedures be investigated. When two or more selection procedures are available which serve a legitimate business purpose with substantially equal validity, the one which has been demonstrated to have less adverse impact should be used.

There are several appropriate methods for determining whether the alternative selection procedures used by agencies are of substantially equal validity as PACE or other tests. One method would be to develop, as part of the required recordkeeping system, a system to track over a period of time the job performance of individuals selected from civil service registers and of individuals selected using alternative procedures. While information resulting from such a tracking system would not scientifically demonstrate the validity of the alternative selection procedures, such a system would provide information on how the procedures compare in terms of the level of job performance. If the tracking system showed that individuals selected on the basis of alternative procedures perform better on the job than individuals selected from registers, consideration could be given to using the alternatives more extensively. If the reverse is shown, then the alternative method should be reexamined to determine whether its use should be continued or the selection procedures for the alternative should be revised to assure validity and job relatedness.

## CONCLUSIONS

There are several alternative selection procedures available to agencies to use in filling vacancies other than selections from OPM job registers. It is generally believed that the alternative procedures have less of an adverse impact on minorities, and OPM informally encourages agencies to use the alternative procedures as a means of achieving their affirmative action goals. However, data has not been maintained to show what proportion of the vacancies in an occupation are filled using the alternative procedures, or whether the alternatives are an effective means of achieving affirmative action goals. Furthermore, neither OPM nor the agencies know which alternative procedure results in selecting employees who perform on the job better or worse than employees obtained from civil service registers.

One way of determining the relative job performance of employees selected using the different procedures would be to establish a system to track over a period of time the job performance of groups of individuals selected using the different procedures. To comply with the Uniform Guidelines, we believe the various alternative selection procedures currently in use by agencies must be investigated to determine if they will achieve the Government's legitimate business purpose with lesser impact than written tests.

## RECOMMENDATIONS

We recommend that the Director, OPM, direct that, as a part of the recordkeeping system required by section 4 of the Uniform Guidelines on Employee Selection Procedures, a system be designed to track over a period of time the job performance of groups of persons selected from civil service registers and those selected using alternative procedures. If the tracking system shows that one or more of the alternatives which have less of an adverse impact than PACE or other tests results in the selection of employees whose job performance is as good as or better than those selected from registers, we then recommend that the Director, OPM, consider directing the expanded use of the alternative procedure. If the tracking system shows that employees selected by an alternative procedure do not perform on the job as well as those selected from registers, we then recommend that the Director reexamine the use of the alternative and determine whether it should be continued.

## CHAPTER 5

### SCOPE OF REVIEW

Our review of the civil service examining and selection program was performed at OPM headquarters in Washington, D.C., and at OPM regional offices in Atlanta, Dallas, Philadelphia, and San Francisco. Our headquarters review included an examination of the extensive research evidence on the development and validation of PACE, and we obtained information about the development of JFA. We also reviewed documentation and discussed the use of unassembled examining procedures.

The data needed to determine the pass rates and racial composition of different applicant groups on the PACE, JFA, Accountant-Auditor examinations, and the Social Worker register was obtained from CSC headquarters, regional and area offices, and the Social Security Administration. The sampling methodology used in making our calculations is described in appendix IV.

We also interviewed and obtained data from (1) personnel management and equal employment opportunity officials of 22 departments and agencies (see app. II), (2) placement officials at a number of universities and colleges around the country, (3) civil rights interest groups, test publishers, and psychologists, and (4) SSA's Office of Research and Statistics. Professional literature on testing was also reviewed.

PACE OCCUPATIONAL COVERAGE LIST

<u>SERIES</u>	<u>TITLE</u>	<u>SERIES</u>	<u>TITLE</u>
011	Bond Sales Promotion	222	Occupational Analysis
018	Safety Management	223	Salary and Wage Administration
020	Community Planning	230	Employee Relations
023	Outdoor Recreation Planner	233	Labor Relations
025	Park Management	235	Employee Development
027	Crop Insurance Administration (except for fieldman and field specialist positions)	244	Labor Management Relations Examining
028	Environmental Protection	246	Contractor Industrial Relations
080	Security Administration	249	Wage and Hour Compliance Specialist
101	Social Science	*301	General Clerical and Administrative
105	Social Insurance Administration	334	Computer Specialist (Trainee)
106	Unemployment Insurance	*341	Administrative Officer
110	Economist	343	Management Analysis
120	Food Assistance Program Specialist	345	Program Analysis
130	Foreign Affairs	346	Logistics Management
131	International Relations	393	Communications Specialist
132	Intelligence	*501	General Accounting Clerical and Administrative
140	Manpower Research and Analysis	526	Tax Technician
142	Manpower Development	560	Budget Administration
150	Geography	570	Financial Institution Examining
170	History	673	Hospital Housekeeping Management
180	Psychology	685	Public Health Program Specialist
184	Sociology		
187	Social Services		
190	General Anthropology	950	Paralegal Specialist
193	Archeology	*962	Contact Representative
201	Personnel Management	965	Land Law Examining
205	Military Personnel Management	967	Passport and Visa Examining
212	Personnel Staffing	987	Tax Law Specialist
221	Position Classification		

<u>SERIES</u>	<u>TITLE</u>	<u>SERIES</u>	<u>TITLE</u>
*990	General Claims Examining	1163	Insurance Examining
991	Worker's Compensation Claims Examining	1165	Loan Specialist
993	Social Insurance Claims Examining	1169	Internal Revenue Officer
994	Unemployment Compensation Claims Examining	1170	Realty
996	Veterans Claims Examining	1171	Appraising and Assessing
997	Civil Service Retirement Claims Examining	1173	Housing Management
		1176	Building Management
		1410	Librarian (for certain trainee positions at GS-5)
		1412	Technical Information
		1420	Archivist
*1001	General Arts and Information (Fine and Applied Arts positions are excluded)	*1421	Archives Specialist
1015	Museum Curator	1654	Printing Management
1081	Public Information	1701	General Education and Training
1082	Writing and Editing	1715	Vocational Rehabilitation
1083	Technical Writing and Editing	1720	Education Research and Program Specialist
*1101	General Business and Industry	1810	General Investigation
1102	Contract and Procurement	1811	Criminal Investigation (except for Treasury Enforcement Agents)
1103	Industrial Property Management	1812	Game Law Enforcement --Covers Marine Enforcement Agents at GS-5 <u>only</u>
1104	Property Disposal		--Does not cover Special Agent (Wildlife)
1130	Public Utility Specialist	1816	Immigration Inspection
1135	Transportation Industry Analysis	1831	Securities Compliance Examining
1140	Trade Specialist	1854	Alcohol, Tobacco, and Firearms Inspection
1145	Agriculture Program Specialist	*1860	Public Health Inspection
1146	Agricultural Marketing		
1147	Agricultural and Fisheries Marketing Reporter		
1149	Wage and Hour Law Administration		
1150	Industrial Specialist		
1160	Financial Analysis		

<u>SERIES</u>	<u>TITLE</u>	<u>SERIES</u>	<u>TITLE</u>
1864	Public Health Quarantine Inspection	2032	Packaging Specialist
1889	Import Specialist	2050	Supply Cataloging
1890	Customs Inspection	*2101	General Transportation
1893	Customs Marine Officer	2111	Transportation Rate and Tariff Examiner
1910	Quality Assurance Specialist	2125	Highway Safety Management
*2001	General Supply	2130	Traffic Management
2003	Supply Program Management	2144	Cargo Scheduling
2010	Inventory Management	*2150	Transport Operations
2030	Distribution Facilities and Storage Management		

This listing is subject to revision as job and qualification requirements are changed.

\*Those series with asterisks are highly varied with respect to job and qualification requirements. Agencies must ensure that the criteria listed below are applied and met for all jobs for which the test is used.

--The position is professional, administrative, or technical in nature (as opposed to clerical or professional support occupations).

--The position has a usual entry grade of GS-5 to GS-7, and a journeyman level of GS-9 or above. If the journeyman level is less than GS-9, the position may be filled through the test if the agency has an established career ladder which provides employees in the position a good opportunity to advance to higher grade level in a related occupational field.

--The abilities measured by the written test are directly related to the position.

Source: U.S. Civil Service Commission Bulletin 337-41, October 16, 1978.

DEPARTMENT AND AGENCY LOCATIONS CONTACTED

Civil Service Commission	Headquarters, Washington, D.C. Atlanta Region, Atlanta Dallas Region, Dallas Philadelphia Region, Philadelphia San Francisco Region, San Francisco Selected area offices in each of these regions and Washington, D.C.
Army Corps of Engineers	San Francisco District, San Francisco
Center for Disease Control (HEW)	Atlanta
Defense Contract Audit Agency	San Francisco Regional Office, San Francisco
Defense Logistics Agency	Headquarters, Washington, D.C. Defense Contract Administration Services Region, Atlanta
Department of the Air Force	Headquarters, Washington, D.C. Warner Robins Air Logistics Center, Robins Air Force Base, Georgia
Department of the Army	Headquarters, Washington, D.C. Forces Command Headquarters, Fort McPherson, Georgia
Department of Health, Education, and Welfare	Headquarters, Washington, D.C. Region III, Philadelphia Region IV, Atlanta Region VI, Dallas
Department of Justice	Headquarters, Washington, D.C.
Department of the Navy	Headquarters, Washington, D.C. Naval Aviation Supply Office, Philadelphia

## APPENDIX II

## APPENDIX II

Department of the Treasury	Headquarters, Washington, D.C.
Equal Employment Opportunity Commission	Region II, Philadelphia Region VII, San Francisco
Federal Aviation Administration (Transportation)	Southwest Regional Office, Fort Worth, Texas
General Services Administration	Headquarters, Washington, D.C.
Internal Revenue Service	Headquarters, Washington, D.C. Mid-Atlantic Region, Philadelphia Southeast Region, Atlanta Southwest Region, Dallas Western Region, San Francisco
National Aeronautics and Space Administration	Ames Research Center, San Francisco
National Park Service (Interior)	Headquarters, Washington, D.C. Mid-Atlantic Region, Philadelphia
U.S. Customs Service (Treasury)	Headquarters, Washington, D.C.
U.S. Fish and Wildlife Service (Interior)	Headquarters, Washington, D.C. Atlanta Region, Atlanta
U.S. Forest Service (Agriculture)	Region V, San Francisco
U.S. Geological Survey (Interior)	Western Region, Menlo Park, California
Veterans' Administration	VA Hospital, San Francisco

PRIOR GAO REPORTS ON THE SUBJECT  
OF EEO IN CIVIL SERVICE EXAMINING AND SELECTION

"Upward Mobility Programs in the Federal Government Should Be Made More Effective," FPCD-75-84, April 29, 1975.

"Progress Made by Agencies in Implementing Upward Mobility Programs," U.S. Civil Service Commission, FPCD-77-10, March 28, 1977.

"Problems of the Federal Employee Equal Employment Opportunity Program Need To Be Resolved," U.S. Civil Service Commission and other Federal agencies, FPCD-76-85, September 9, 1977.

"Conflicting Congressional Policies: Veterans' Preference and Apportionment Vs. Equal Employment Opportunity," FPCD-77-61, September 29, 1977.

"Problems with Federal Equal Employment Opportunity Guidelines on Employee Selection Procedures Need To Be Resolved," FPCD-77-54, February 2, 1978.

SAMPLING TECHNIQUES

The following tables describe how well individuals performed on PACE during fiscal year 1976 by race, region, proportion of competitors, in various score ranges, augmented and raw scores, outstanding scholar and quality graduate status. We used a CSC duplicate tape of fiscal year 1976 PACE competitors to determine the total number of qualified GS-5 competitors in each stratum. All GS-5 applicants were studied in the Atlanta, Dallas, Philadelphia, and San Francisco CSC regions as separate stratum, and in the remaining CSC regional offices combined into one stratum. Persons with invalid social security numbers (less than 1 percent) were eliminated from the universe counts.

Once universe counts were established, simple random sampling techniques were used to calculate sampling sizes for a maximum sampling error of 10 percent at the 95-percent level of confidence. Nationwide results were arrived at by appropriate stratified sampling techniques. The Social Security Administration provided us with data on the racial composition--percentage white, black, other races, and unknown--of the competitors in each stratum.

Simple random sampling techniques used to select sample sizes for a maximum sampling error of 10 percent at the 95-percent level of confidence were also used to sample JFA, accountant/auditor, and social worker applicants.

## PACE--AUGMENTED SCORES

Test score strata	Universe	Sample	White		Black		Other		Unknown	
			Percent	Sampling error (note a)						
<u>Atlanta Region</u>										
90+	2,728	106	89.6	5.7	0.0	0.0	0.0	0.0	10.4	5.7
70-89	9,197	109	79.8	7.5	7.3	4.9	0.0	0.0	12.8	6.3
69-	13,805	109	75.2	8.1	18.3	7.3	0.0	0.0	6.4	4.6
Total	25,730	324	78.4	5.2	12.5	4.3	0.0	0.0	9.1	3.4
Pass rate (note b)			<u>48.5</u>		<u>21.0</u>		<u>0.0</u>		<u>62.3</u>	
Certi fiable rate (note b)			<u>12.1</u>		<u>0.0</u>		<u>0.0</u>		<u>12.0</u>	
<u>Dallas Region</u>										
90+	1,551	102	92.2	5.1	1.0	1.9	0.0	0.0	6.9	4.8
70-89	5,260	108	88.0	6.1	3.7	3.5	0.9	1.8	7.4	4.9
69-	8,364	109	77.1	7.9	16.5	7.0	0.9	1.8	5.5	4.3
Total	15,175	319	82.4	4.9	10.5	4.0	0.8	1.2	6.3	2.9
Pass rate (note b)			<u>48.4</u>		<u>13.2</u>		<u>38.9</u>		<u>51.9</u>	
Certi fiable rate (note b)			<u>11.4</u>		<u>.1</u>		<u>0.0</u>		<u>11.1</u>	
<u>Philadelphia Region</u>										
90+	2,895	105	87.6	6.2	1.0	1.8	0.0	0.0	11.4	6.0
70-89	8,542	109	87.2	6.3	3.7	3.5	0.0	0.0	9.2	5.4
69-	10,465	109	76.1	8.0	15.6	6.8	1.8	2.5	6.4	4.6
Total	21,902	323	82.0	4.6	9.0	3.5	0.9	1.2	8.2	3.1
Pass rate (note b)			<u>55.6</u>		<u>17.3</u>		<u>0.0</u>		<u>62.4</u>	
Certi fiable rate (note b)			<u>14.1</u>		<u>1.4</u>		<u>0.0</u>		<u>18.5</u>	

a/If the sampling error is larger than the percent, then the lower limit of the confidence interval is greater than zero because at least one case was found in the sample.

b/Pass and certifiable rates were calculated using estimated numbers for the strata.

Test score strata	Universe	Sample	White		Black		Other		Unknown	
			Percent	Sampling error (note a)						
<u>San Francisco Region</u>										
90+	3,548	107	84.1	6.9	0.0	0.0	4.7	4.0	11.2	5.9
70-89	8,300	109	78.9	7.6	3.7	3.5	6.4	4.6	11.0	5.9
69-	8,811	109	67.9	8.8	20.2	7.5	8.3	5.2	3.7	3.5
Total	20,659	325	75.1	5.0	10.1	3.5	6.9	3.0	7.9	3.0
Pass rate (note b)			<u>61.4</u>		<u>14.6</u>		<u>49.0</u>		<u>80.2</u>	
Certifiable rate (note b)			<u>19.2</u>		<u>0.0</u>		<u>11.6</u>		<u>24.3</u>	
<u>All Other Regions</u>										
90+	15,293	109	86.2	6.5	0.0	0.0	1.8	2.5	11.9	6.1
70-89	40,464	110	87.3	6.2	1.8	2.5	1.8	2.5	9.1	5.4
69-	44,245	110	70.0	8.6	20.9	7.6	.9	1.8	8.2	5.1
Total	100,002	329	79.5	4.7	1.0	3.5	1.4	1.3	9.1	3.3
Pass rate			<u>61.0</u>		<u>7.4</u>		<u>71.7</u>		<u>60.3</u>	
Certifiable rate			<u>16.6</u>		<u>0.0</u>		<u>19.8</u>		<u>20.0</u>	
<u>Nationwide</u>										
90+	26,015	529	86.8	4.0	.2	.2	1.7	1.6	11.3	3.8
70-89	71,763	545	85.4	3.9	3.1	1.7	1.8	1.5	9.7	3.3
69-	85,690	546	72.1	4.9	19.3	4.3	1.6	1.1	7.0	2.9
Total	183,468	1,620	79.4	2.8	10.3	2.1	1.7	.8	8.6	1.9
Pass rate			<u>57.6</u>		<u>12.0</u>		<u>55.8</u>		<u>62.4</u>	
Certifiable rate			<u>15.5</u>		<u>0.2</u>		<u>14.1</u>		<u>18.6</u>	

a/If the sampling error is larger than the percent, then the lower limit of the confidence interval is greater than zero because at least one case was found in the sample.

b/Pass and certifiable rates were calculated using estimated numbers for the strata.

Test score strata	PACE--RAW SCORES									
	Universe	Sample	White		Black		Other		Unknown	
			Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)
<u>Atlanta Region</u>										
90+	1,545	103	88.3	6.0	0.0	0.0	0.0	0.0	11.7	6.0
70-89	8,195	109	91.7	5.2	1.8	2.5	0.0	0.0	6.4	4.6
69-	15,990	109	69.7	8.6	22.9	7.9	0.0	0.0	7.3	4.9
Total	25,730	321	77.9	5.6	14.8	5.0	0.0	0.0	7.3	3.4
Pass rate (note b)			<u>44.3</u>		<u>3.9</u>		<u>0.0</u>		<u>37.6</u>	
<u>Dallas Region</u>										
90+	829	96	85.4	6.7	0.0	0.0	0.0	0.0	14.6	6.7
70-89	4,720	107	95.3	4.0	1.9	2.5	0.0	0.0	2.8	3.1
69-	9,626	109	80.7	7.4	14.7	6.6	0.0	0.0	4.6	3.9
Total	15,175	312	85.5	4.9	9.9	4.3	0.0	0.0	4.6	2.7
Pass rate (note b)			<u>40.1</u>		<u>5.9</u>		<u>0.0</u>		<u>36.4</u>	
<u>Philadelphia Region</u>										
90+	1,830	104	79.8	7.5	0.0	0.0	2.0	2.6	18.3	7.2
70-89	8,214	109	90.8	5.4	2.8	3.1	0.9	1.8	5.5	4.3
69-	11,858	109	75.2	8.1	17.4	7.1	1.8	2.5	5.5	4.3
Total	21,902	322	81.5	4.9	10.5	4.0	1.5	1.5	6.6	2.9
Pass rate (note b)			<u>50.0</u>		<u>9.9</u>		<u>33.5</u>		<u>54.6</u>	

a/If the sampling error is larger than the percent, then the lower limit of the confidence interval is greater than zero because at least one case was found in the sample.

b/Pass rates were calculated using estimated numbers for the strata.

Test score strata	Universe	Sample	White		Black		Other		Unknown	
			Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)
<u>San Francisco Region</u>										
90+	2,269	105	77.1	7.9	0.0	0.0	7.6	5.0	15.2	6.7
70-89	8,037	109	83.5	7.0	1.8	2.5	9.2	5.4	5.5	4.5
69-	10,353	109	69.7	8.6	13.8	6.5	11.9	6.1	4.6	3.9
Total	20,659	323	75.9	5.2	7.6	3.4	10.4	3.7	6.1	2.7
Pass rate (note b)			<u>54.0</u>		<u>9.4</u>		<u>42.4</u>		<u>62.4</u>	
<u>All Other Regions</u>										
90+	9,843	109	85.3	6.6	0.9	1.8	0.9	1.8	12.8	6.3
70-89	39,093	110	88.2	6.1	0.0	0.0	4.5	3.9	7.3	4.9
69-	51,066	110	72.7	8.4	12.7	6.3	4.5	3.9	10.0	5.6
Total	100,002	329	80.0	4.9	6.6	3.2	4.2	2.5	9.2	3.5
Pass rate (note b)			<u>53.6</u>		<u>1.4</u>		<u>44.6</u>		<u>44.6</u>	
<u>Nationwide</u>										
90+	16,316	517	83.9	4.3	0.6	1.1	1.8	1.3	13.8	4.0
70-89	68,259	544	88.9	3.7	0.9	0.6	3.8	2.3	6.4	2.9
69-	98,893	546	73.0	4.8	15.2	3.7	3.8	2.1	7.9	3.1
Total	183,468	1,607	79.9	2.9	8.6	2.0	3.6	1.4	7.9	2.0
Pass rate (note b)			<u>50.7</u>		<u>4.4</u>		<u>43.3</u>		<u>45.8</u>	

a/If the sampling error is larger than the percent, then the lower limit of the confidence interval is greater than zero because at least one case was found in the sample.

b/Pass rates were calculated using estimated numbers for the strata.

PACE OUTSTANDING SCHOLARS'AUGMENTED SCORES

Test score strata	Universe	Sample	White		Black		Other		Unknown	
			Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)
<u>Atlanta Region</u>										
90+	1,407	101	88.1	6.1	1.0	1.9	0.0	0.0	10.9	5.9
70-89	2,048	103	59.2	9.3	35.9	9.1	0.0	0.0	4.9	4.1
69-	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	3,455	204	71.0	6.0	21.7	5.4	0.0	0.0	7.3	3.4
Certifiable rate (note b)			<u>50.6</u>		<u>1.9</u>		<u>0.0</u>		<u>60.7</u>	
<u>Dallas Region</u>										
90+	781	95	94.7	4.2	0.0	0.0	0.0	0.0	5.3	4.2
70-89	1,046	100	72.0	8.4	22.0	7.8	0.0	0.0	6.0	4.4
69-	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	1,827	195	81.7	5.1	12.6	4.4	0.0	0.0	5.7	3.1
Certifiable rate (note b)			<u>49.6</u>		<u>0.0</u>		<u>0.0</u>		<u>39.4</u>	
<u>Philadelphia Region</u>										
90+	1,406	102	92.2	5.0	0.0	0.0	0.0	0.0	7.8	5.0
70-89	1,498	102	79.4	7.6	12.7	6.3	0.0	0.0	7.8	5.1
69-	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	2,904	204	85.6	4.6	6.6	3.2	0.0	0.0	7.8	3.6
Certifiable rate (note b)			<u>52.1</u>		<u>0.0</u>		<u>0.0</u>		<u>48.5</u>	

a/If the sampling error is larger than the percent, then the lower limit of the confidence interval is greater than zero because at least one case was found in the sample.

b/Certifiable rates were calculated using estimated numbers for the strata.

Test score strata	Universe	Sample	White		Black		Other		Unknown	
			Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)
<u>San Francisco Region</u>										
90+	1,610	102	91.2	5.4	0.0	0.0	1.0	1.9	7.8	5.1
70-89	1,325	102	83.3	7.0	4.9	4.0	6.9	4.7	4.9	4.0
69-	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	2,935	204	87.6	4.3	2.2	1.8	3.6	2.4	6.5	3.3
Certifiable rate (note b)			<u>57.1</u>		<u>0.0</u>		<u>15.0</u>		<u>66.0</u>	
<u>All Other Regions</u>										
90+	7,500	108	83.3	7.0	0.0	0.0	0.0	0.0	16.7	7.0
70-89	7,684	108	73.1	8.3	12.0	6.1	3.7	3.6	11.1	5.9
69-	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	15,184	216	78.2	5.5	6.1	3.1	1.9	1.8	13.9	4.6
Certifiable rate (note b)			<u>52.6</u>		<u>0.0</u>		<u>0.0</u>		<u>59.4</u>	
<u>Nationwide</u>										
90+	12,704	508	86.5	4.3	0.1	0.2	0.1	0.2	13.2	4.3
70-89	13,601	515	72.6	5.1	15.8	3.8	2.8	2.1	8.8	3.5
69-	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	26,305	1,023	79.4	3.3	8.2	2.0	1.5	1.1	10.9	2.7
Certifiable rate (note b)			<u>52.7</u>		<u>0.6</u>		<u>4.1</u>		<u>58.4</u>	

a/If the sampling error is larger than the percent, then the lower limit of the confidence interval is greater than zero because at least one case was found in the sample.

b/Certifiable rates were calculated using estimated numbers for the strata.

<u>PACE OUTSTANDING SCHOLARS</u>										
<u>RAW SCORES</u>										
<u>Test score strata</u>	<u>Universe</u>	<u>Sample</u>	<u>White</u>		<u>Black</u>		<u>Other</u>		<u>Unknown</u>	
			<u>Percent</u>	<u>Sampling error (note a)</u>						
<u>Atlanta Region</u>										
90+	492	88	88.6	6.0	0.0	0.0	0.0	0.0	11.4	6.0
70-89	1,483	102	87.3	6.3	2.9	3.2	0.0	0.0	9.8	5.6
69-	1,480	102	57.8	9.3	39.2	9.2	1.0	1.9	2.0	2.6
Total	3,455	292	74.8	4.9	18.1	4.2	0.4	0.8	6.7	2.8
Pass rate (note b)			<u>66.9</u>		<u>7.1</u>		<u>0.0</u>		<u>87.4</u>	
<u>Dallas Region</u>										
90+	265	77	84.4	6.9	1.3	2.1	1.3	2.1	13.0	6.4
70-89	831	97	91.7	5.2	0.0	0.0	0.0	0.0	8.2	5.2
69-	731	95	66.3	8.9	24.2	8.1	3.2	3.3	6.3	4.6
Total	1,827	269	80.5	4.4	9.9	3.2	1.4	1.4	8.2	3.1
Pass rate (note b)			<u>67.0</u>		<u>1.7</u>		<u>11.5</u>		<u>69.1</u>	
<u>Philadelphia Region</u>										
90+	508	90	84.4	6.8	1.1	2.0	0.0	0.0	14.4	6.6
70-89	1,420	101	84.2	6.9	2.0	2.6	0.0	0.0	13.9	6.5
69-	976	99	64.6	9.0	26.3	8.3	1.0	1.9	8.1	5.1
Total	2,904	290	77.7	4.7	10.0	3.1	0.3	1.6	12.0	3.8
Pass rate (note b)			<u>72.0</u>		<u>11.7</u>		<u>0.0</u>		<u>77.4</u>	

a/If the sampling is larger than the percent, then the lower limit of the confidence interval is greater than zero because at least one case was found in the sample.

b/Pass rates were calculated using estimated numbers for the strata.

Test score strata	Universe	Sample	White		Black		Other		Unknown	
			Percent	Sampling error (note a)						
<u>San Francisco Region</u>										
90+	680	95	89.5	5.8	0.0	0.0	2.1	2.7	8.4	5.2
70-89	1,376	101	92.1	5.1	1.0	1.9	1.0	1.9	5.9	4.5
69-	879	97	75.3	8.1	6.2	4.5	13.4	6.4	5.2	4.2
Total	2,935	293	86.4	3.7	2.3	1.6	5.0	2.2	6.3	2.7
Pass rate (note b)			<u>73.9</u>		<u>20.6</u>		<u>19.2</u>		<u>75.5</u>	
<u>All Other Regions</u>										
90+	3,005	105	82.9	7.1	1.0	1.8	0.0	0.0	16.2	7.0
70-89	7,386	108	91.7	5.2	1.0	1.8	0.9	1.8	6.5	4.6
69-	4,793	107	73.8	8.3	13.1	6.3	5.6	4.3	7.5	5.0
Total	15,184	320	84.3	3.9	4.8	2.2	2.2	1.6	8.7	3.1
Pass rate (note b)			<u>72.3</u>		<u>13.4</u>		<u>20.2</u>		<u>73.0</u>	
<u>Nationwide</u>										
90+	4,950	455	84.6	4.5	1.0	1.1	0.3	0.4	14.3	4.4
70-89	12,496	509	90.3	3.3	1.2	1.2	0.7	1.1	7.8	3.0
69-	8,859	500	69.7	5.0	19.1	4.6	4.9	2.5	6.3	2.8
Total	26,305	1,464	82.3	2.4	7.2	1.5	2.0	1.0	8.5	1.9
Pass rate (note b)			<u>71.5</u>		<u>10.2</u>		<u>18.5</u>		<u>75.1</u>	

a/If the sampling is larger than the percent, then the lower limit of the confidence interval is greater than zero because at least one case was found in the sample.

b/Pass rates were calculated using estimated numbers for the strata.

JUNIOR FEDERAL ASSISTANT EXAMINATION

PASS AND CERTIFIABLE RATES

Strata (note c)	Universe	Sample	White		Black		Other		Unknown	
			Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)
<u>Atlanta Region</u>										
Certifiable	43	43	79.1	0.0	16.3	0.0	4.7	0.0	0.0	0.0
Failed	32	32	50.0	0.0	50.0	0.0	0.0	0.0	0.0	0.0
Total	75	75	66.7	0.0	30.7	0.0	2.7	0.0	0.0	0.0
Pass rate (note b)			<u>68.0</u>		<u>30.4</u>		<u>100.0</u>		<u>0.0</u>	
Certifiable rate (note b)			<u>68.0</u>		<u>30.4</u>		<u>100.0</u>		<u>0.0</u>	
<u>Dallas Region</u>										
Certifiable	13	13	76.9	0.0	15.4	0.0	7.7	0.0	0.0	0.0
Others passed	267	78	83.3	7.0	12.8	6.3	0.0	0.0	3.8	3.6
Failed	152	63	52.4	9.5	39.7	9.3	6.3	4.6	1.6	2.4
Total	432	154	72.2	5.5	22.2	5.1	2.5	1.6	2.8	2.4
Pass rate (note b)			<u>74.4</u>		<u>37.5</u>		<u>9.1</u>		<u>83.3</u>	
Certifiable rate (note b)			<u>3.2</u>		<u>2.1</u>		<u>9.1</u>		<u>0.0</u>	
<u>Philadelphia Region</u>										
Certifiable	113	113	74.3	0.0	20.4	0.0	0.9	0.0	5.3	0.0
Others passed	282	282	47.2	0.0	47.9	0.0	0.0	0.0	5.0	0.0
Failed	465	465	17.4	0.0	80.6	0.0	0.0	0.0	1.3	0.0
Total	860	860	34.7	0.0	62.0	0.0	0.1	0.0	3.4	0.0
Pass rate (note b)			<u>72.8</u>		<u>29.6</u>		<u>100.0</u>		<u>69.0</u>	
Certifiable rate (note b)			<u>28.2</u>		<u>4.3</u>		<u>100.0</u>		<u>20.7</u>	

a/If the sampling error is larger than the percent, then the lower limit of the confidence interval is greater than zero because at least one case was found in the sample.

b/Pass and certifiable rates were calculated using estimated numbers for the strata.

c/Certifiable scores vary from time to time and from region to region depending upon job market conditions. At the time of our review certifiable scores ranged from 70 in the Atlanta region to 90 in the Dallas region.

Strata (note c)	Universe	Sample	White		Black		Other		Unknown	
			Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)
<u>San Francisco Region</u>										
Certifiable	36	36	83.3	0.0	8.3	0.0	5.6	0.0	2.8	0.0
Others passed	343	343	81.3	0.0	10.2	0.0	4.7	0.0	3.8	0.0
Failed	185	185	49.2	0.0	41.1	0.0	8.6	0.0	1.1	0.0
Total	564	564	70.9	0.0	20.2	0.0	6.0	0.0	2.8	0.0
Pass rate (note b)			<u>77.3</u>		<u>33.3</u>		<u>52.9</u>		<u>87.5</u>	
Certifiable rate (note b)			<u>7.5</u>		<u>2.6</u>		<u>5.9</u>		<u>6.3</u>	
			<u>Total white</u>		<u>Total black</u>		<u>Total other</u>		<u>Total unknown</u>	
Certifiable	205	205	77.1	0.0	17.1	0.0	2.9	0.0	3.4	0.0
Others passed	892	703	71.1	2.1	22.9	1.9	1.8	0.0	4.1	1.1
Failed	834	745	32.1	1.7	63.2	1.7	3.1	0.8	1.2	0.4
Total	1,931	1,653	54.9	1.2	39.7	1.1	2.5	0.4	2.8	0.5
Pass rate (note b)			<u>74.7</u>		<u>31.2</u>		<u>45.8</u>		<u>81.5</u>	
Certifiable rate (note b)			<u>14.9</u>		<u>4.6</u>		<u>12.5</u>		<u>13.0</u>	

a/If the sampling error is larger than the percent, then the lower limit of the confidence interval is greater than zero because at least one case was found in the sample.

b/Pass and certifiable rates were calculated using estimated numbers for the strata.

c/Certifiable scores vary from time to time and from region to region depending upon job market conditions. At the time of our review certifiable scores ranged from 70 in the Atlanta region to 90 in the Dallas region.

ACCOUNTANT-AUDITOR CERTIFIABLE RATES

Strata	Universe	Sample	White		Black		Other		Unknown	
			Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)	Percent	Sampling error (note a)
<u>Atlanta Region</u>										
Certifiable	153	153	87.6	0.0	5.9	0.0	1.3	0.0	5.2	0.0
Not certifiable	149	149	86.6	0.0	8.1	0.0	0.0	0.0	5.4	0.0
Total	302	302	87.1	0.0	7.0	0.0	0.7	0.0	5.3	0.0
Certifiable rate (note b)			<u>51.0</u>		<u>42.9</u>		<u>100.0</u>		<u>50.0</u>	
<u>Dallas Region</u>										
Certifiable	277	277	85.2	0.0	6.5	0.0	1.8	0.0	6.5	0.0
Not certifiable	674	674	83.2	0.0	9.2	0.0	1.2	0.0	6.4	0.0
Total	951	951	83.8	0.0	8.4	0.0	1.4	0.0	6.4	0.0
Certifiable rate (note b)			<u>29.6</u>		<u>22.5</u>		<u>38.5</u>		<u>29.5</u>	
<u>Philadelphia Region</u>										
Certifiable	140	140	89.3	0.0	2.9	0.0	0.0	0.0	7.9	0.0
Not certifiable	226	226	87.6	0.0	4.9	0.0	0.9	0.0	6.6	0.0
Total	366	366	88.3	0.0	4.1	0.0	0.5	0.0	7.1	0.0
Certifiable rate (note b)			<u>38.7</u>		<u>26.7</u>		<u>0.0</u>		<u>42.3</u>	
<u>San Francisco Region</u>										
Certifiable	267	267	83.1	0.0	3.0	0.0	7.9	0.0	6.0	0.0
Not certifiable	901	138	80.4	6.1	0.0	0.0	11.6	0.0	8.0	4.2
Total	1,168	405	81.1	4.7	0.7	0.0	10.7	0.0	7.5	3.2
Certifiable rate (note b)			<u>23.4</u>		<u>100.0</u>		<u>16.8</u>			
<u>Total white</u>										
Certifiable	837	837	85.7	0.0	4.7	0.0	3.3	0.0	6.3	0.0
Not certifiable	1,950	1,187	82.7	2.8	4.4	0.0	5.8	2.3	7.1	1.9
Total	2,787	2,024	83.6	2.0	4.4	0.0	5.1	1.6	6.9	1.3
Certifiable rate (note b)			<u>30.8</u>		<u>31.5</u>		<u>19.7</u>		<u>27.7</u>	

a/If the sampling error is larger than the percent, then the lower limit of the confidence interval is greater than zero because at least one case was found in the sample.

b/Certifiable rates were calculated using estimated numbers for the strata. Applicants who meet minimum qualification standards were considered to have passed the examination. Certifiable scores vary from time to time and from region to region depending upon job market conditions.

COMPENDIUM OF PUBLICATIONS BY THEOFFICE OF PERSONNEL MANAGEMENTON THE DEVELOPMENT AND RESEARCH OF PACE

McKillip, R. H., Trattner, M. H., Corts, D. B., and Wing, H.  
The Professional and Administrative Career Examination:  
Research and Development. Washington, D.C.; U.S. Civil  
Service Commission, Personnel Research and Development  
Center, Personnel Research Report 77-1, 1977.

## Authors' Abstract

The work of the United States Civil Service Commission's Personnel Research and Development Center to develop a basis for the written test portion of the Professional and Administrative Career Examination (PACE) is described. The objective of the research was the identification of a construct valid set of cognitive abilities tests appropriate for selection of applicants for entry level positions in Federal professional, administrative, and technical occupations. The occupational coverage of the examination was defined and a set of 27 occupations selected for intensive study. These occupations represented approximately 70 percent of annual appointments in the occupations to be covered by the PACE. Duties of the 27 occupations were rated by subject matter experts for importance and relative amount of time spent in their performance. Subject matter experts also rated a specially developed set of 31 knowledges, skills, abilities, and other worker characteristics (KSAO's) in terms of their importance for overall job performance. Six cognitive abilities were hypothesized as important for duty performance and were rated by personnel research psychologists. A method was devised for combining subject matter expert ratings with psychologist ratings to determine the relative weight of each ability within each occupation. Patterns of ability weights to be applied to subtests for each job resulted from this process. Test question types were identified from the professional literature as measures of the abilities to be included in the test. Factor analysis of the subject matter expert rating of the 31

KSAO's provided support for the abilities.  
 Research needed to provide additional technical  
 support for the test was outlined.

Reck, M. Criterion-Related Validity of PACE Constructs: A  
 Review of the Literature. Washington, D.C.; U.S. Civil  
 Service Commission, Personnel Research and Development  
 Center, Professional Series 77-4, 1977.

#### Authors' Abstract

The Federal Government uses the PACE written  
 examination to select employees for large number  
 of jobs in many professional, technical and admin-  
 istrative occupations. This paper reviews studies  
 of criterion-related validity of constructs meas-  
 ured by the PACE for jobs similar or identical  
 to those for which selections are made through  
 that examination. Given the reported validity of  
 a large number of tests measuring PACE constructs  
 in a variety of settings, it is concluded that  
 these constructs are generally related to satisfac-  
 tory performance on jobs similar or identical  
 to those for which the PACE examination is used.

Northrop, L. C. The Definition and Measurement of Judgment.  
 Washington, D.C.; U.S. Civil Service Commission, Personnel  
 Research and Development Center, Technical Memorandum  
 76-17, 1976.

#### Authors' Abstract

The complex nature of the psychological con-  
 struct called judgment, the history of its iden-  
 tification as a mental ability, its measurement,  
 and its relationship to other mental abilities  
 are presented and discussed in this paper. Fac-  
 tor analytic research which isolated the judgment  
 factor is reviewed and instruments used to measure  
 it are examined.

Northrop, L. C. The Definition and Measurement of Numerical  
 Ability. Washington, D.C.; U.S. Civil Service Commission,  
 Personnel Research and Development Center, Technical Memo-  
 randum 77-9, 1977.

#### Authors' Abstract

This report traces the factor-analytic his-  
 tory of the isolation and identification of the

cognitive ability construct known as Number or Numerical Facility. Empirical documentation from the psychometric literature is presented and discussed and examples of items from a variety of tests which have defined and been associated with the Number factor are included to illustrate the type of tasks which require this ability.

Northrop, L. C. The Definition and Measurement of Verbal Comprehension. Washington, D.C.; U.S. Civil Service Commission, Personnel Research and Development Center, Technical Memorandum 77-11, 1977.

#### Authors' Abstract

The factor-analytic history of the isolation and identification of the mental ability construct known as Verbal Comprehension is traced through the psychometric literature in this report. Examples of the most important item types which have been found to mark this factor are presented and discussed. The association of some of these item types with other ability factors is also pointed out.

Northrop, L. C. The Definition and Measurement of Reasoning. Washington, D.C.; U.S. Civil Service Commission, Personnel Research and Development Center, Technical Memorandum 77-10, 1977.

#### Authors' Abstract

The complex and multi-faceted nature of the cognitive process of reasoning is discussed in this report. The factor-analytic history of the isolation and definition of the three most important currently recognized reasoning factors (General Reasoning, Induction, and Logical or Deductive Reasoning) is traced through the psychometric literature and the interdependence of these abilities is discussed. Examples of many test item-types which have been found to define these factors are presented with empirical documentation of their importance and use in a great variety of test batteries and with different examinee populations.

O'Leary, B. S. and Trattner, M. H. Research Base for the Written Test Portion of the Professional and Administrative Career Examination (PACE): Prediction of Job Performance for Internal Revenue Officers. Washington, D.C.; U.S. Civil Service Commission, Personnel Research and Development Center, Technical Study 77-6, 1977.

#### Authors' Abstract

This study was one of a number of criterion-related validity studies conducted with Test 500, the written test portion of the Professional and Administrative Career Examination. A concurrent validity model was followed with a very high use PACE occupation, Internal Revenue Officer. Journeyman employees took Test 500, a job task inventory, a biographical information blank, a job information test, and a work sample test. The job task inventory furnished information utilized in the construction and scoring of all the criterion instruments. Additionally, revenue officers were evaluated by their first-level supervisors. Test 500 scores were significantly correlated with all criteria.

Trattner, M. H., Corts, D. B., van Rijn, P. P., and Outerbridge, A. M. Research Base for the Written Portion of the Professional and Administrative Career Examination (PACE): Prediction of Job Performance for Claims Authorizers in the Social Insurance Claims Examining Occupation. Washington, D.C.; U.S. Civil Service Commission, Personnel Research and Development Center, Technical Study 77-3, 1977.

#### Authors' abstract

This study was the first in a series of criterion-related validity studies conducted with Test 500, the written test portion of the Professional and Administrative Career Examination. A concurrent validity model was followed with a very high-use PACE occupation, Social Security Administration Claims Authorizer. Journeyman employees took Test 500, a job task inventory, a biographical information blank, a job information test, and a work sample test. The job task inventory furnished information utilized in the construction and scoring of all the criterion

instruments. Additionally, claims authorizers were evaluated by their first-level supervisor. The results revealed that Test 500 scores correlated highly and significantly with all criteria.

O'Leary, B. S. Research Base for the Written Portion of the Professional and Administration Career Examination (PACE): Prediction of Training Success for Social Insurance Claims Examiners. Washington, D.C.; U.S. Civil Service Commission, Personnel Research and Development Center, Technical Study 77-5, 1977.

#### Authors' Abstract

This study was one of a number of studies conducted with Test 500, the written test portion of the Professional and Administrative Career Examination. Specifically, the present study was aimed at determining the criterion-related validity of Test 500 in the prediction of training performance for Social Insurance Claims Examiners in the Social Security Administration. Total Test 500 scores, when weighted according to three weighting methods (construct ability weight, equal weight, and optimal weight) correlated highly and significantly with average training performance as measured by training tests given throughout a 12-week training program.

Corts, D. B., Muldrow, T. W. and Outerbridge, A. M. Research Base for the Written Test Portion of the Professional and Administrative Career Examination (PACE): Prediction of Job Performance for Customs Inspectors. Washington, D.C.; U.S. Civil Service Commission, Personnel Research and Development Center, Technical Study 77-4, 1977.

#### Authors' Abstract

As a part of the overall research plan in support of the construct validity of Test 500, which is the written test portion of the Professional and Administrative Career Examination (PACE), a series of concurrent criterion-related validity studies was carried out. These studies were to determine (a) whether statistically significant relationships exist between an ability subtest-weighted Test 500 and specially developed job performance

criteria; (b) whether the ability weighting system specially devised for Test 500 was statistically as effective as other methods commonly used for weighting; and (c) other statistical relationships between Test 500, job performance, and biographical variables. Essentially, this third study in the series evaluates the effectiveness of the ability-weighted Test 500 as a part of the ranking process for selecting applicants into the occupation of Customs Inspector. Research instruments were developed in cooperation with subject matter experts, and consisted of a job task inventory, a biographical information blank, a job information test, a work sample, and supervisory rating and ranking forms. Job task inventory data formed the basis for construction and scoring of the criterion instruments. Customs inspectors were rated and ranked by their supervisors. Journeyman-level research participants numbered 190 from 14 sites nationwide. Test 500 was found to be significantly related to job performance as reflected in job information test and work sample scores. No significant relationships were found between Test 500 and supervisory ratings and rankings, probably as a result of problems in the supervisory appraisal data. The operational ability-weighted Test 500 was found to be statistically as effective as equal weighting and as effective as optimal weighting based upon multiple correlation with the job information test and the work sample. The results of this study provide further support for the construct validity of Test 500.

Martin, C. G. and Pavlov, Y. Preliminary Estimates of Means, Standard Deviations, and Reliabilities for Test 500 of the Professional and Administrative Career Examination (PACE). Washington, D.C.; U.S. Civil Service Commission, Personnel Research and Development Center, Technical Memorandum 75-2, 1975.

#### Authors' Abstract

This memorandum shows how item statistics, i.e., item difficulties (p's) and item-test correlations (r's), collected on varying experimental groups and test forms can be used to estimate means, average p's and r's, standard

deviations, and reliabilities of an operational test. The equations used were derived from relationships between item and test statistics for a single group taking a single test form. Estimated test statistics for eight series of Test 500 of the Professional and Administrative Career Examination are presented.

Martin, C. G. A Procedure for Estimating the Reliability of a Weighted Linear Composite Test: The Reliability of Test 500, Series 110. Washington, D.C.; U.S. Civil Commission, Personnel Research and Development Center, Technical Memorandum 75-4, 1975.

#### Authors' Abstract

A research design is presented and discussed for estimating the reliability of a weighted linear composite when only one test form is administered and when alternate forms of subtests are administered. Subtest, total test and pattern total score reliabilities for both cases are presented for Test 500, Series 110 of the Professional and Administrative Career Examination (PACE). Associated common scale score means, variances and standard deviations are also presented.

AGENCY COMMENTS

OPM has primary responsibility for developing and administering the examination programs discussed in this report. At our request, it formally commented on the draft report.

EEOC, the Department of Justice, the U.S. Commission on Civil Rights, the Merit System Protection Board, the Office of Education in the Department of Health, Education, and Welfare also have specific interests and/or responsibilities for the matters discussed in this report. They were also given an opportunity to comment.

EEOC and the U.S. Commission on Civil Rights provided formal comments. The Department of Justice declined to comment because it is representing the Government in a lawsuit (Luevano, et al. v. Campbell, No. 79-0271, D. D. C. filed Jan. 29, 1979) which alleges that PACE is discriminatory and violates the Civil Rights Act of 1964. The Merit System Protection Board and the Office of Education did not make any comments on the report.

OPM's comments indicated that it believes that PACE is a fully validated test for efficiently examining a very large number of applicants and that the frequently voiced criticisms of the examination have little merit. Nevertheless, it indicated general agreement with the report recommendations but made no specific commitment to implement the recommendations.

EEOC said that the report appropriately applied the principles of the Uniform Guidelines on Employee Selection Procedures but did not go far enough in its conclusions and recommendations. EEOC believes the current method for using PACE should be discontinued, at least until more research work is performed. The U.S. Commission on Civil Rights said that regardless of the validity of PACE, it screens out too many minorities and should be replaced with a less discriminatory alternative.

The Chairwoman, Subcommittee on Civil Service, House Committee on Post Office and Civil Service, has scheduled hearings for May 15, 1979, to explore the various views regarding the use of examinations such as PACE in the Government. The Chairwoman has requested that we issue this report in time for the hearings.

Because of the hearings schedule, we did not have time to analyze, follow up, evaluate, and discuss the agencies' comments fully in the body of the report as is our usual practice. However, to the extent time permitted, we considered the agencies' comments in the final report. We also added our note at the end of each agency's comments where we believed additional commentary was necessary.

United States of America  
**Office of  
Personnel Management**

Washington, D C 20415

In Reply, Refer To

Your Reference

**APR 13 1979**

Mr. H. L. Krieger  
Director, Federal Personnel  
and Compensation Division  
General Accounting Office  
Washington, D.C. 20548

Dear Mr. Krieger:

This will forward for your consideration the comments of the professional staff in the Office of Personnel Management (OPM) on the General Accounting Office (GAO) draft report entitled: "Federal Employment Examining and Selecting Procedures: Do They Achieve Equal Opportunity and Merit Principle Goals?"

The report deals with a highly difficult, technical, and sensitive issue. The auditors are to be commended for their extensive study of the technical as well as programmatic issues. In our view, the body of the report generally reflects reasonably informed understanding of these issues. The staff has, however, identified certain errors of fact and conclusion which we believe you will want to correct. The enclosed staff comments also include some additional information which we believe should be reflected in the final report.

We want to draw your particular attention to the major problems presented by the executive summary. In contrast to the reasonably objective treatment of the issues in the body of the report, the summary deals with the facts in a way that may be seen as undermining the report's objectivity. This is especially evident in the repeated and uncritical reference to criticisms of the PACE examination by other agencies without any evaluation of the adequacy of documentation available to support these criticisms or the countervailing point of view of the OPM professional staff. For these reasons, our staff believes that a redraft of the summary is necessary to assure the credibility of the report.

In view of the significance of the PACE in the Federal examining program and the disproportionate attention given to PACE in the GAO report in contrast to the other examinations reviewed, we believe it important that the following points be emphasized in the summary as well as the report proper:

1. The PACE, on the basis of OPM's professional staff analysis, is a fully validated instrument for efficiently examining a very large number of applicants (135,000 in FY 1978) from whom high quality candidates are selected for important entry-level positions.

2. The adverse impact of PACE with regard to minority applicants is not a phenomenon peculiar to this examination, but rather, is typical of instruments used for measuring the same kinds of abilities as those important to success in PACE-type occupations. The search for an alternative measurement device of equal validity but little or no adverse impact has a desirable objective but one not easy to attain.
  
3. The Civil Service Commission and the Office of Personnel Management have long recognized that entry into PACE-type jobs should not be restricted to a single examining device, and have in fact encouraged widespread use of other means of entry such as promotion, upward mobility, Veterans Readjustment Appointments, cooperative education, and other educationally related programs. These alternatives have greatly ameliorated adverse impact by bringing women and minorities into the Federal work force in sufficient numbers to satisfy the "bottom line" concept sanctioned in the Uniform Guidelines. Nevertheless we remain concerned that OPM's PACE examination appears to have such a high degree of adverse impact. Particularly because it is so extensively used in hiring recent college graduates it is incumbent upon us to continually examine the feasibility of alternatives. Historically, high proportions of people hired through this type of examination later achieve high rank in the career service. Given the congressional mandate in the Civil Service Reform Act to "achieve a workforce from all segments of society," we must continue to explore the use of alternative methods. We are working with the Equal Employment Opportunity Commission and the Department of Justice on this issue.

We appreciate the opportunity to comment on this most important report. We are concerned, as we know you are, that a factual and fair representation of the issues be presented to the Congress. With this in mind, we would welcome further discussion at the staff level on any of our suggested changes to the report. If you find this appropriate after reviewing our comments, arrangements may be made with Arch Ramsay, our Deputy Associate Director for Staffing, on 632-6005. At the same time you have my assurances that we will continue to reexamine critically any selection procedure which shows a high degree of adverse impact.

Sincerely yours,

  
 Alan K. Campbell  
 Director

Enclosure

Comments by OPM Professional Staff  
on the GAO Draft Report, "Federal Employment  
Examining and Selecting Procedures: Do They  
Achieve Equal Opportunity and Merit Principle  
Goals?"

SECTION I: CRITICAL PROBLEMS IN THE REPORT

1. P. iii and P. 33. [ii & 21]

Minority Representation Among Incumbents is not Assigned Its Proper Role.

GAO states it "did not reach a conclusion concerning the question of adverse impact," then makes observations about the minority makeup of the workforce in PACE occupations. It makes the point that minority group members comprise about 17 percent of incumbents in GS 5-11 for populous PACE occupations, but offers no evaluation of these data or their importance. GAO then states that the impact of a selection device is based on applicants rather than incumbents. What GAO ignores is the first part of the title of its report, i.e., do Federal selection procedures achieve equal opportunity? If equal opportunity is measured by reasonableness of minority representation in the workforce, adverse impact of individual selection devices becomes secondary.

While it is true that UGL's focus adverse impact determinations on the applicant group, the discussion in the GAO Report gives the misleading impression that the minority representation among incumbents is irrelevant. Actually, the UGL's state that the affirmative action records and accomplishments of an employer are relevant and have to be considered in any EEO determinations (see Section 4E of UGL's). The unusually high minority representation in PACE occupations is therefore obviously a critical consideration, and the report should note this fact. (See related comment No. 15.)

2. P. iv, para. 3 and summary as a whole. [iii]

Summary Wrongly Emphasizes the Negative on PACE Validation Studies.

GAO asks three questions about validity but addresses only the third question in the summary, i.e., whether the tests were "properly validated." The text, in discussing PACE validity, provides answers to the first two questions [is there a clear relationship between test and job performance and do tests predict job performance? (See pp. 48 and 49)]. Yet no conclusions related to these questions are provided in the summary.

This is one example of a general deficiency in the summary (pp. i through xi). While the text discusses both the positive and negative aspects of the PACE validation effort as seen by GAO, the summary focuses almost exclusively on the negative judgments. For example, with respect to the three questions raised there (at the bottom of p. iv), the text provides affirmative answers to two of them:

- a. The text shows there is a clear relationship between job performance and test scores (see especially the graphs on p. 48).

-2-

- b. The text shows test scores do predict job performance (see p. 49), [35] yet the summary focuses only on GAO's doubts as to whether the PACE has been "properly validated" (see pp. v ff.).

That is, the summary focuses only on the answer to the third question, and in so doing does not accurately represent the contents or full import of the report. This point is particularly critical, since many individuals may have time to read only the summary. We strongly recommend the summary be rewritten to be an accurate reflection of the text.

3. P. v and p. 50. [iv & 38]

Mention Should be Made of Technical Expertise of Criticizing Agencies.

In the interests of fairness and accuracy, this discussion should point out the relative professional and technical resources available in the agencies criticizing PACE and in OPM. While the OPM staff includes a large number of experienced selection psychologists, to our knowledge no such professional and technical expertise is available in the CRC, the Civil Rights Task Force of the President's Reorganization Project, or in the Civil Rights Division of Justice. This is a highly relevant item of information to the reader attempting to evaluate technical comments on PACE and its associated studies emanating from these agencies.

GAO cites only critics from special interest organizations who have often fought against use of standardized tests and have never supported use of such tests. These groups lack the technical expertise to evaluate validity evidence. GAO did not report (and it appears did not even seek) the evaluation of competent professional groups or other disinterested parties. This gives the appearance that GAO did not perform an objective, unbiased review of Federal examining.

4. P. vi. [iv]

Uniform Guidelines Were Not in Effect When PACE Validation Studies Took Place.

Here and throughout the report, GAO states that the UGL's must be followed. However, these guidelines were not in effect when these examinations in question were developed and the UGL's permit use of reports that met previous guidelines [see Section 15A(3)(6) of the UGL's]. PACE research certainly meets the guidelines in effect at the time it was conducted. Further, PACE meets all UGL's requirements for interim use of a selection procedure; this includes substantial existing evidence of validity (which GAO cites in the report) and a study (underway) to meet any new guidelines requirements not already met. These facts should be included in the summary and in the text.

-3-

5. P. vi, Point 2. [iv & 41]

Job Analysis Information Inaccurate.

GAO cites critic's complaints that job analyses "were conducted for only 27 occupations." This is not accurate. Job analyses were conducted for all 118 occupations. The job analyses of 27 occupations in which the great majority of placements were made constituted only one phase of the research. The remaining occupations were studied separately. Apart from this inaccuracy, GAO does not indicate why it believes more occupations should be analyzed, how many more and which ones. The initial PACE research sampled about 23% of the occupations, covering about 70% of the hires. Later job analyses added the remainder of the total of 118 occupations to the PACE coverage, but also resulted in deletion of a number of occupations which had been covered by the predecessor examination.

6. P. vi, Points 3 and 4. [v]

Number of Criterion-Related Studies; Predictive vs. Concurrent Design.

There were actually four criterion-related validity studies on the PACE --not three (see also our comment on pp. 45 ff). We believe these studies fully support PACE use under the construct validation strategy. Does GAO believe further studies are needed, and if so, why and how many?  
[See GAO note 1, p. 111.]

While the studies performed to date have been of the concurrent rather than predictive design for reasons cited in the draft report, we do plan a research study of the type suggested on applicants selected under an alternative procedure (see related comment no. 11).

7. P. vii, Points 5, 6, and 7. [v & vi]

Collection of Racial and Ethnic Data to Assure That Research Participants Are Representative of Relevant Labor Market; Investigation of Test Un-Fairness; Consideration of Alternatives to PACE.

OPM has continued over a number of years to investigate alternatives with potentially less adverse impact even without determining whether they were of substantially equal validity. We agree that additional consideration of alternatives having less adverse impact is in order (see related comment no. 11).

8. P. viii and pp. 62 and 63. [vi & 50]

New JFA Examination Development Does Not Reflect a Belief That Current JFA is Not Valid.

There is a misunderstanding of the facts in this recommendation and discussion concerning the JFA. The JFA is currently used for 28 occupations. There are more minorities and women represented in these occupations than

-4-

in the overall workforce. For these 28 occupations, validity data were collected on 13; no attempt was made to locate data for the other 15 because the data on these 13 more than supported the examination's use for all of the 28 occupations since the occupations have similar ability requirements. The work that OPM has underway to develop a new examination is part of the overall validation work going on for all examinations in OPM and does not reflect any doubt that the JFA examination is valid for these occupations.

Three reports have been prepared on the current written test for the JFA. The first one shows that all forms are reliable and meet appropriate psychometric construction considerations. The second one shows the support in the psychometric literature for the particular item types used in the test. The third summarizes the relevant criterion-related validation studies for 13 of the 28 occupations used in this program. These reports are available from OPM and should be cited and summarized in the GAO report. Their omission would create serious distortion.

9. P. x and pp. 14 and 78, Recommendation. [iii, vii, x, & 22]

Race/Ethnicity Data Collection Efforts at OPM Have Been Underway Since 1976.

The recommendation that records be maintained on applicants by race and ethnicity is relevant. However, the report should note that CSC started immediately upon the signing of the Federal Executive Agency Guidelines to set up its records keeping systems for this work. Since that time, extensive experimental studies have been conducted to carry out this work and there is a program to set up the operational aspects of this system in progress at the present time. However, there are three major problems that OPM faces in this area.

- a. Funding. Race and ethnic data collection and tracking is extremely expensive. Where we have in place automated examination procedures which include certification histories, such tracking can be done for those modes of entry covered by the system. Where we do not have automated certification historical records kept--the usual case--the tracking of applicants from the point of hire is prohibitively expensive and can be done for only a few examinations at a time. If a tracking system should be set up for all modes of entry in order to capture the bottom line data as is required by the UGL's, the costs would run into the tens of millions of dollars and would need to be far more thorough than our current CPDF systems of tracking for current Federal employees. In the CPDF we have no records concerning how people were considered for any position nor how many people were considered for the positions. In order to capture bottom line data, a matrix approach of capturing all data for each position would need to be established and systems for combining these data into job categories would have to be made. For the present time, OPM plans to collect these data on high priority examinations which

-5-

have a likelihood for high adverse impact and on high volume examinations rather than collect these data for all examinations.

- b. Response rate to requests for self-identification of race and sex by questionnaire. Based on our research, there is a very good response rate where written tests are used (95 percent of the applicants who appear in the test rooms complete the forms). As a result, OPM is prepared to collect data on those examinations where written tests are used. However, where written tests are not used in the examination process, the return rate is about 60 percent. This is far too low to produce meaningful results.
- c. Forms clearance problems. While OMB gave OPM clearance to collect data during FY 77 and 78 on applicants to try out different procedures and to set up our systems, it has not yet given clearance for FY79 and beyond, nor has it cleared a general form for other Federal agencies to use. Based on the FY 77 and 78 data, however, OPM is putting in place the procedures to collect these data routinely in the future as funds and clearances become available.

In the interests of accuracy, the report should note all of these facts.

10. P. xi, para. 1. [x]

Logic on Validation of Alternative Procedures is Unclear.

If the alternate procedure shows a useful level of validity, then ordinarily maximum validity and job-relatedness would be obtained by combining the alternate procedure with the existing valid procedure (PACE), i.e., using both valid procedures. Also, it is not clear what is meant by the "opposite" outcome. But if this term is meant to include selecting employees with substantially lower performance on the job than those selected from the register, what is called for is not validation but abandonment of the procedure.

In considering alternative procedures, it is important to recognize that the most common feature of PACE occupations (Appendix 1 - pp. 80-81) is their information burden. Many jobs not labeled as "Analysis," "Investigation," "Examining," or "Inspection" positions relate either to a specific profession (General Schedule 110, 150-193, 950) or to topic areas for which the label "specialist" implies mastery of a body of knowledge. There exists a complex written body of laws and regulations which affects each PACE job. The cognitive abilities measured by PACE are, therefore, clearly related to success in these occupations and it would be illogical to assume that ways may be found to bypass consideration of these abilities and still to ensure competence in these occupations.

This fact--that PACE jobs are information-handling jobs--is a bedrock feature of the selection problems relating to PACE. Although there are additional job-related skills not measured by the PACE due to feasibility

-6-

problems (character variables, personality variables and skills mentioned such as long-term memory), the necessity for paying close attention to information skills (the "constructs" measured by Test 500 clearly relate to the intake, evaluation and processing of information) cannot be ignored. There is no known cost-effective means of measuring them other than with PACE or a similar instrument. We strongly recommend that the report recognize these realities.

11. P. xi and p. 69, Recommendation. [x & 55]

GAO Recommendation for Predictive Research Study, Assuring to Extent Possible That Participants are Representative of the Labor Market, and for Investigation of Test Fairness.

Although predictive studies are not necessary to establish validity, OPM is currently planning a study with the Social Security Administration in which "all applicants [will be] selected for a PACE occupation by some other merit procedure and [will be] administered the test and followed up at a later date" to determine relative job performance. In the course of this study we will, to the extent feasible, assure that participants are representative of the relevant labor market and investigate test fairness.

12. Pp. 17-28. [9-15 & 23]

Since GAO recommendations for further validation effort on PACE are based in the magnitude of adverse impact, it is important to recognize that the GAO survey findings may for several reasons exaggerate racial differences in PACE scores:

- a. Competition from private employers for the most able black college graduates, including offers of premium salaries, may have resulted in an underrepresentation of this vital group among those taking the PACE. If this group typically does not take the PACE, this fact would account for part of the test score difference in the GAO data.
- b. Our research indicates that people obtaining lower scores on PACE are more likely to re-take the examination quickly. Since blacks average lower scores as a group, it is possible that proportionately more blacks—especially those with lower scores—have been PACE repeaters than in the case of whites.
- c. The four regions selected for study are not representative of the nationwide PACE population. The two southern regions of Dallas and Atlanta consistently show the lowest average PACE scores. (See p. 85.) [72]

In this connection, the report should also take note of the extensive evidence that adverse impact is not unique to PACE, i.e., the fact that it has been the usual finding among college graduate populations similar to PACE competitors. It occurs, for example, in tests used to

-7-

determine qualification for entry into professional schools. This fact is essential to an accurate understanding of Federal selection problems because it makes clear that tests like PACE do not create adverse impact. The adverse impact exists long before candidates take the test; the test merely reveals previous real educational deficiencies.

13. P. 28, ff. [17]

Adverse Impact Not Determined for Unassembled Examinations.

The basic question of adverse impact on applicants is not answered for the Accountant-Auditor unassembled examination. Pass rates and "certifiable rates" are unknown for both blacks and whites. All that is known is that of those who did pass, comparable percentages of blacks and whites were in the ranges of scores where certification to an agency is likely. Therefore, no conclusions about adverse impact on applicants can be drawn. Elsewhere the report insists adverse impact must be assessed on applicants (see page 33). The same considerations apply in the case of the Social Worker unassembled examination. Here again, neither pass rates nor "certifiable" rates are known for either race for applicants. Again, therefore, adverse impact on applicants cannot be determined. The broad conclusionary statements on pages iii and 34 that "blacks scored proportionately as well as whites on the unassembled examinations" are therefore not justified by the evidence. These statements should be modified to be consistent with the available data.

14. P. 31, last para. [22]

Report Omits Mention of Regulations Which Previously Restrained OPM From Race Data Collection.

Here we find a statement that the reason information is not generally available to allow determination of adverse impact is "inadequate record-keeping." Actually, until recently, we were prohibited from gathering racial data on applicants to Federal jobs. On p. 33 (last paragraph) the report implies that OPM has not acted since the rule change on the problem of collecting race data. In fact, OPM has devoted substantial resources to developing a workable and accurate system for collecting such data. For example, a study has been conducted to determine the accuracy of questionnaire self reports of race. Although the administrative and other problems are quite difficult, some such data have already been collected and are being analyzed. (See also our related comment No. 9.)

15. Pp. 35-36. [24]

Report Discounts "Bottom Line." No Reasons Provided For Labeling PACE "Unusual."

For the reasons given in 1, above, it is clear that PACE is indeed a "usual" circumstance. It is, in fact, an unusually usual circumstance. Because of the high minority representation in PACE occupations, PACE is

clearly an example of a case in which enforcement action would be inappropriate even though one component of the selection procedure does have adverse impact. This is clearly the kind of situation contemplated by the writers of the UGL's (see Section 3C of the UGL's and Q&A No. 25 on the UGL's; Section 4E is also relevant.). It is also the same set of facts faced by the Supreme Court in Davis. The report, to be accurate, should note that the 17% minority representation is even higher relative to relevant workforce figures and to private employers representation in comparable occupations than it is relative to population figures. No reasoning to the contrary is offered in the report. The report simply states, without providing reasons, that "PACE cannot be considered to be a usual circumstance," (p. 36, para. 1). (24)

With respect to the recommendation (p. 36) on data collection, we do plan to collect minority applicant data on written tests once funding and forms clearance problems are resolved (see related comment 9). It would be impossible, however, to break down the PACE data by occupation, since the examination is used to fill 118 different occupations, for most of which all successful candidates are eligible.

[See GAO note 5, p. 111.]

17. P. 44, para. 1. [30]

Omission of SME Participation.

GAO should point out that 1241 subject matter experts from many agencies and geographical locations participated in the development of PACE.

[See GAO note 5, p. 111.]

19. P. 45, ff. [32]

Research Participants Underreported.

GAO has omitted the criterion-related study in the Bureau of Disability Insurance of SSA which was provided. This study, added to the others, brings the research participant sample to 901. [See GAO note 1, p.111.]

20. Pp. 48-49. [36]

OPM Research on Productivity and Selection Should be Mentioned.

Omitting any mention of this area of research on the PACE is difficult to understand, especially in light of:

- a. the current national emphasis on the importance of increasing productivity, especially in the Federal government;
- b. the attention the report gives to other facets of the PACE research program; and
- c. the findings that PACE has a substantial impact on workforce output.

This information was made available to GAO and in fact was included in earlier versions of this report. We strongly recommend that it be restored to the final GAO report.

21. P. 53, last para. and its continuation on p. 54. [40]

Uniform Guidelines Requirements Misrepresented.

This seems to misrepresent UGL's requirements. A description of the research sample and how it was selected are essential documentation requirements, but its comparison with the relevant labor market is desirable, not essential. The samples used in the PACE criterion-related validity studies were thoroughly described. [See GAO note 2, p. 111.]

22. P. 51. [41]

Some FSEE Occupations Dropped from PACE Due to Job Analysis.

Although the original PACE R&D work was based on 27 occupations, the remaining occupations were included only if they survived an appropriate job analysis. Indeed, those analyses resulted in a number of occupations, covered by FSEE, being dropped from PACE. (See similar statement in our comment No. 5.)

23. P. 52. [39 & 40]

Serious Contradiction on Construct Validity History.

Here the report quotes the UGL's to the effect that "construct validity is a relatively new and developing procedure in the employment field." On p. 53, the report states that construct validity was the basis for military testing in World Wars I and II and is at present the basis for the Armed Services Vocational Aptitude Battery used by all branches of the military. Further, it is the basis of testing done by the U.S. Employment Service in civilian employment. Clearly, this is a contradiction which should be resolved in favor of the historical facts.

24. P. 54, para 1. [40]

Adverse Impact does not Negate Validity Evidence.

The report states that:

In our opinion, if the test had less of an adverse impact, or was less important in obtaining a professional or an administrative job in the civil service, then the validation strategy and evidence supporting the test would be very persuasive. We believe, however, that in view of the impact the test has on black applicants, and its importance in obtaining a civil service job, each of the professional standards and Uniform Guidelines requirements must be scrupulously followed.

Either the validity evidence is sound or it is not. If it is, the test meets the UGL's. There is no provision for double standards based on degree of adverse impact. Also, as indicated earlier, GAO's sampling procedures may overestimate degree of adverse impact. (See our comment No. 12.)

25. P. 54, para. 2. [41 & 42]

Validity Generalization/Inaccurate Job Analysis Information.

OPM now has additional studies underway for creating a further basis for generalizing across jobs based on common duties. Also as indicated earlier

(see comment No. 5), job analysis was performed on all 118 occupations--not just the 27 core occupations.

26. Pp. 57-60 and 67-68. [43 to 45 & 56]

Predictive vs. Concurrent Studies.

This section on predictive vs. concurrent validity studies contains a number of errors:

- [44] a. It implies (see line 13, p. 56) that in predictive studies applicants are selected at random. This is in fact never done. Employees are selected either on the test being validated or on another correlated procedure. Thus, contrary to the report, predictive studies rarely "permit the analysis of the job performance of persons who do not score well on written tests" (p. 56, paragraph 1). That is, problems of restricted test score ranges are, in general, just as frequent in predictive as concurrent studies. Similarly, restriction in range resulting from turnover for failure to perform adequately or from accelerated promotion of exceptional employees will be present in both predictive and concurrent studies.
- b. The facts in (a) above explain empirical findings that contradict an important implication in the report. The report holds that the distinction between predictive and concurrent studies is important because the two kinds of studies often--perhaps typically--produce different results. In fact, the empirical evidence indicates just the opposite: the two kinds of studies typically provide very similar results in the case of aptitude or ability tests like PACE or the JFA. This fact was pointed out by Lee Cronbach in 1970; we have since found the same thing to be true in our validity generalization research. The conclusion holds in educational as well as employment testing. Results obtained in validity studies depend on the general methodological quality of the study (e.g., sample size, reliability of job performance measures), not on whether the study is predictive or concurrent. Since conclusions about test validity do not depend on this distinction, there is no basis for the report's conclusion that predictive studies are essential.
- [44] c. On p. 56, paragraph 2, GAO indicates a belief that concurrent studies tend to underestimate validity. If so, GAO apparently must conclude that (1) existing estimates of PACE validity (which are substantial) are underestimates and (2) it is not particularly important to conduct predictive validity studies, since these would only produce still higher validity figures.

27. Pp. 67-68. [45-48]

Inaccurate and Contradictory Treatment of Test Fairness Issues.

This section contains important logical contradictions. The report states correctly at the top of p. 59 that research has shown that tests valid

for whites are valid for blacks. On the next page, the report states, again correctly, that research shows that "tests are either fair to minority groups or slightly biased in their favor" by the definition of test fairness which "maximizes productivity and is based on competition or merit and equal opportunity for the individual" (p. 59).

Strangely, the report does not endorse this concept of selection fairness as the appropriate one--despite the fact that it is the only concept consistent with the merit system requirements imposed by law, requirements which are recognized elsewhere in the report. Further, it is the only concept of selection or test fairness ever endorsed by a court of law (*Cortez v. Rosen*, Circuit Court decision, California). A change is clearly needed here.

The research findings cited in the report make it clear that:

- a. Representativeness of the study samples in terms of race is not critical or even relevant in determining test validity. Validities have been shown to be the same for blacks and whites; and
- b. Test fairness studies are not critical, since the research literature indicates that valid tests are almost always fair to blacks (by the only definition of fairness legally compatible with a merit system).

Ironically, the report then goes on to criticize PACE research on grounds that:

- a. Samples used have not been shown to be representative (especially with respect to race).
- b. Test fairness studies have not been conducted.

An analysis of available sample sizes in PACE validity studies showed that fairness studies were not technically feasible. [This is likely to remain true in any future PACE validity studies.] In connection with the topic of fairness, the report should note that the test construction process each item was carefully reviewed for possible bias before inclusion in the test. Thus, OPM did look into questions of bias or unfairness. The report should note both these facts.

28. P. 76, para 2. [62]

Validity of Alternative Procedures.

Here we find the statement that "Neither OPM nor the agencies know whether the alternative selection procedures result in hiring individuals who perform better or worse on the job than individuals selected from civil service registers." This statement conflicts with the presentation in the [35] report (p. 49) of research evidence from three occupations showing conclusively that test-selected employees show higher performance on a variety

of different performance measures than employees selected using alternate procedures. Further, this difference is quite substantial in magnitude. Thus, solid information bearing on this question is available and is in fact included in GAO's report. Further, in light of what is known from the research literature on the validity of a wide variety of selection methods, there is every reason to expect that future studies will produce similar results.

In this connection, the report should note another potential problem with alternatives to PACE. It is possible that such alternatives would lead to fragmentation of entry-level professional selection into 118 separate, non-comparable systems. Some implications of such a development are:

- a. Applicants may have to apply separately for each occupation in which they have an interest, a heavy burden;
- b. Increased credentialism and specialization and consequent increased difficulty of occupational entry for young and inexperienced applicants; and
- c. A great increase in administrative (and validation) costs.  
[See GAO note 3, p. 111.]

[See GAO note 5, p. 111.]

#### SECTION II: IMPORTANT BUT NOT CRITICAL PROBLEMS IN THE REPORT

##### 30. P. i, para. 3. [i]

Inadequate Treatment of Representativeness of Study Samples.

The report fails to state how or why the four examinations studied were determined to be representative of OPM examinations, or even whether they were determined to be representative. In light of the heavy emphasis the report places on the importance of sample representativeness, this omission is hard to justify. Elsewhere the report implies that unless sample representativeness is assured, conclusions drawn are open to question.

##### 31. P. i, para. 3. [i]

Improper Terminology: "Test" vs. "Examination."

GAO refers to PACE and JFA "tests," but to Accountant-Auditor and Social Worker "examinations." PACE and JFA are not tests; they are examinations that

include written cognitive ability tests and other measurements (tests) such as of education and experience, academic achievement, and veteran preference. Accountant-Auditor and Social Worker examinations also include tests of education, experience, and veteran preference. A similar confusion of terms occurs on p. ii, where in the first full paragraph reference is made to unassembled tests.

32. P. ii. [i]

Consolidation of Major Questions Needed.

It would be helpful if GAO would state its principal questions early in the summary section so as to establish for the reader the approach followed. The first full paragraph lists the "first" question GAO asked, but not until p. iv is another question asked.

33. P. viii. [vi]

No Reference Made to Job-Relatedness of Unassembled Examinations.

GAO did not review the job-relatedness of unassembled examinations. This omission is not consistent with the title of the report and the statement on p. i that GAO wanted to determine whether examining procedures are achieving the policy goal of selection on the basis of relative ability, knowledge, and skills. (See related comment no. 46.) Actually, we have been concerned for some time with with a need to improve the job-relatedness of unassembled examinations. This concern has led to the development of the new methodology based upon the behavioral consistency model which is referred to in the GAO report, and to the development of interim procedures to improve the job-relatedness of rating schedules used for "deferred rating" type examinations.

34. P. 4. [3]

Inaccuracies with Regard to Scoring of Tests.

The description of scoring is not quite accurate. Applicants don't earn scores of "at least 70 points (on a scale of 100)." They earn scores period; and if the scores are high enough they are converted to ratings, ranging between 70 and 100.

Also, compensably disabled veterans not only receive 10 extra points if they are eligible (pass the written test), they are placed at the top of the register for most jobs at GS-9 and below. That is, they are placed above people with higher scores, even with the extra added points, and CP vets cannot be passed over without very good reason.

35. P. 8, para. 2. [5]

District Court Decision Summary Needed.

In the discussion of Douglas v. Hampton, the District Court decision as well as the Appeals Court decision should be summarized in the interests of accuracy.

36. P. 8, 1st para. [5]

Language Change.

The reference to "minority or female" is erroneous. The reference should be to a "race, sex, or ethnic group."

37. P. 10, bottom para. [6]

Incomplete Reference to Uniform Guidelines.

Omits third option—otherwise demonstrate business necessity. (Cf. Section 6B of UGL's.)

38. P. 18, last para. [11]

Transmuted vs. Raw Scores.

Raw scores do not vary only between 40 and 100. They vary between 0 and 150. GAO is writing about transmuted scores.

39. P. 24. [15]

Percentages Differ.

The percentages of Outstanding Scholars who were black and who were white do not agree with the percentages reported on p. 22. [13]  
[See GAO note 4, p. 111.]

40. P. 35, 1st para. [24]

Language Changed.

Relevant comparison is relevant labor force, not overall labor force.

41. P. 39. [27]

Construct Validity Misdefined.

The definition of construct validity is misleading. Constructs are not "constructed from a theory of how people behave." Rather, theories of human behavior are built up from the results of studies and observations, and constructs are components of these theories. Further, it is now widely recognized that constructs are as "observable"

as other forms of behavior. Most constructs--including those making up PACE--are operationally defined by the procedures used to measure them.

42. P. 42. [29]

Reference to Validity Generalization is Out of Date.

The argument concerning limited generalizability of criterion-related studies does not accurately reflect the reasoning of the PACE researchers or recent developments in research on validity generalization. If criterion-related validity were the sole strategy, the studies would not serve to explicate the methodology of the job analysis, the procedures used in test specification and construction, or the scientific judgment required to relate the written test to the job analysis. Recent research in validity generalization has shown that test validity is much more generalizable than previously thought. Validity generalization, criterion-related validity, and construct validity are all interactive aspects of PACE validity.

43. P. 47, para 2. [32]

Constraints on Use of Supervisory Rankings Should be Mentioned.

The report should note that the reasons supervisor rankings of job performance were not used in the internal revenue officer study was that the officers' union contract forbade such usage.

44. Page 47, last line. [33]

Language Improved.

The term physical separation is too vague. The report should state explicitly that supervisors of customs inspectors did not have adequate opportunities to observe subordinate performance.

45. P. 56a. Drawbacks of Suggested Methodology Should be Mentioned. [43]

The report makes suggestions as to how predictive validity studies can be made technically feasible. These suggestions will be workable only if the numbers hired onto a specific job over a reasonable period of time (1-2 years) will be adequate. In most cases, the numbers will be too small and scattered (among locations) to permit technical feasibility. Also, any valid selection procedure will create test score range restriction, as indicated above. Nevertheless, as indicated in our comment No. 11, OPM is currently planning a study of the sort the report recommends here.

46. Pp. 63-67 [51-53]

Omission.

The report does not address unassembled examining in relation to UGL's requirements.

47. P. 76, para. [61]

Error Regarding Alternate Methods.

Some "alternate methods" of selection do require competitive examination. Candidates for the DOD cooperative education program must score at least 70 on the PACE to be selected.

GAO notes: To the extent time permitted, we considered and discussed OPM's comments in this final report. However, because of the timeframe involved for publishing the final report, we did not discuss or respond to each comment in the body of the report. Where we believed additional commentary was needed but not specifically provided in the report, we provided the following GAO notes.

1. The fourth study referred to by OPM concerned the prediction of training success in one of the three occupations studied. While not specifically discussed in our report, it was considered in our evaluation of the PACE validation strategy.
2. The Uniform Guidelines requirements are not misrepresented. A description of the race, sex, and ethnic composition of the sample is an essential standard. GAO did not describe the comparison with the relevant labor market as essential, although as OPM observed, it is very desirable.
3. There is no conflict in the report statement referred to by OPM. OPM's research, from which the table was derived, did not identify performance levels by individual alternative but rather, all the alternatives were combined. It is possible that each alternative results in employing a different quality of individual, some better than those selected from the register, and some worse.
4. The difference referred to is due to a statistical sampling error.
5. Deleted comments relate to matters discussed in the draft report but deleted in the final report.
6. Page numbers in brackets refer to pages in this report.



EQUAL EMPLOYMENT OPPORTUNITY COMMISSION  
WASHINGTON, D. C. 20506

OFFICE OF THE CHAIR

April 16, 1979

Gregory J. Ahart, Director  
Human Resources Division  
United States General Accounting Office  
Washington, D.C. 20548

Dear Mr. Ahart:

This is in reply to your letter of March 28, 1979 transmitting a draft of a proposed GAO staff report entitled "Federal Employment Examining and Selecting Procedures: Do They Achieve Equal Opportunity and Merit Principle Goals?"

This report shows an excellent grasp of the basic principles of the Uniform Guidelines on Employee Selection Procedures and appears, generally, to do a good job of applying those principles to the selection procedures studied. However, there are several additional conclusions and important recommendations which flow logically from the material presented but which are not adequately stressed in the document as drafted.

Part I of this letter summarizes the major elements of the Guidelines which are reflected in the draft report. Part II notes some elements of the Guidelines which the draft report does not adequately cover. Part III suggests additional conclusions which GAO should consider adding that flow from its own factual findings. Part IV comments on the recommendations which GAO did make in the draft. Part V suggests that the most important single failing of OPM's approach to selection procedures is the failure to adequately seek alternative methods of selecting and urges that GAO add additional recommendations to its report to deal with this situation. Finally, Part VI deals with a number of smaller items which relate to the points made in the first five parts of the letter.

The material in Part V contains our most important recommendation to GAO.

It should be pointed out that we have not had access to the studies submitted by CSC which were examined by GAO.

I. Draft report reflects basic principles of the Uniform Selection Guidelines:

The GAO staff draft report on federal examining clearly presents major principles and requirements of the Uniform Guidelines on Employee Selection Procedures as they apply to all employers including the Federal Government as an employer. The following principles were mentioned in the document and should remain in the final version:

- A. The basic principle of the Guidelines, which is the basic principle of Federal anti-discrimination law, states that an employment selection procedure which has an adverse impact is illegal if it cannot be justified.
- B. Adverse impact is defined by the Uniform Selection Guidelines as a substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group.
- C. Validated selection procedures, in instances in which there is adverse impact, are those which have been validated in accord with the technical standards of the Uniform Selection Guidelines including an investigation of suitable alternatives.
- D. There are three validity strategies which are appropriate for demonstrating the validity of a selection procedure: criterion-related validity, content validity, and construct validity.
- E. All three validation strategies attempt to demonstrate whether or not a selection procedure provides significant information about probable future job performance. Thus, a job analysis is a critical part of a validity study.
- F. The Uniform Selection Guidelines require that validation studies include an investigation of suitable alternative selection procedures and also suitable alternative methods of using the selection procedures. Such investigation is crucial for eliminating "artificial, arbitrary, and unnecessary barriers to employment".
- G. The Uniform Selection Guidelines also recognize the need for allocating federal agency resources to maximize enforcement impact on employers who are excluding people from hire, promotion, and training on prohibited grounds. This is the "bottom line" approach to allocation of administrative and prosecutorial resources. EEOC has further spelled out this enforcement emphasis in its Standards for Systemic Cases.

II. Requirements of the Uniform Selection Guidelines omitted from the draft report:

The GAO staff draft report failed to identify five important requirements of the Uniform Selection Guidelines:

A. Interim use:

The use of selection procedures which have an adverse impact and are not yet validated may not lawfully be continued unless the Uniform Selection Guideline requirements of Section 5J on interim use have been met. GAO does not address this issue.

B. Employer has two options:

The basic principle of the Guidelines contained in sections 3A and 6A is that an employer utilizing a procedure with an adverse impact has two options. The draft report fails to make clear these two options and thus its recommendations tend to focus on only one of the options -- validation. The two options are:

1. to validate the selection procedure; or
2. to use alternative selection procedures which facilitate the hiring of qualified workers while eliminating adverse impact.

C. "Bottom Line" is not a rule of law:

In the Uniform Selection Guidelines the federal enforcement agencies adopted a "bottom line" approach only for the exercise of their administrative and prosecutorial discretion in usual circumstances. The "bottom line" approach was not adopted as a rule of law. See question #26 of the Questions and Answers on the Uniform Selection Guidelines, 44 FR 11996 (Friday, March 2, 1979). Thus, GAO is not precluded by the "bottom line" approach from finding that the PACE and JFA examinations have adverse impact on blacks. (See paragraphs III A and B below)

D. Adverse impact must be determined for each group:

The Uniform Selection Guidelines require that each race, sex, and ethnic group be considered by itself. These groups are specified in Section 4B of the Guidelines. Impact on groups may not be calculated together as "minorities" or "other minorities." Thus, the GAO cannot make a finding concerning the impact of the examinations on Hispanics, Asians, or American Indians, under the grouping of "other minorities." (See: paragraph III C, below.) [See GAO note 1, p. 124.]

E. Validation evidence must be sufficient to support the method of use of the selection procedure:

Section 5G of the Uniform Selection Guidelines requires that if the use of a selection procedure for ranking has greater adverse impact than its use on an appropriate pass/fail basis, then the evidence of validity and utility must be sufficient to support its use on a ranking basis. Technical requirements are specified in Section 14B (5) and (6). Although GAO found that certifiable scores have a greater disproportionate impact on blacks than do pass/fail scores, GAO did not address the question of sufficiency of evidence for use of scores on a ranking basis. [See GAO note 2, p. 124.]

III. Suggested additional conclusions for the GAO report based on GAO findings:

- A. The GAO report finds that blacks pass and receive certifiable scores on the PACE and the JFA examinations at a substantially lower rate than whites. Applying the definition of adverse impact, Section 16B, to these passing and certifiable rates, GAO could conclude that both tests have an adverse impact on blacks. The statement of the Assistant Attorney General, Civil Rights Division, Department of Justice, quoted on page 50 of the GAO Draft Report, supports this conclusion. As pointed out in IIC above, this conclusion is not prohibited by the "bottom line" approach of Section 4C.
- B. The GAO report finds that OPM and its predecessor agency, the Civil Service Commission, failed to collect, maintain, and have available the data on adverse impact. GAO points out that this data is not only required by the Uniform Selection Guidelines issued in 1978, but was also required by the Federal Executive Agency Guidelines on Selection Procedures adopted by the Civil Service Commission in November of 1976. The GAO could also add that such data not only continues to be unavailable but that instructions have been issued to other federal agencies not to collect such data pending instructions from OPM on how to do so. (See Federal Personnel Management Letter 300-25.) From these findings, GAO could conclude that even if OPM chooses to defend its use of the PACE and JFA examinations on grounds that the "bottom line" of the total selection processes is "good", it is prohibited from doing so at this time by its own failure to collect the relevant "adverse impact" information.

- C. GAO found information on race and ethnicity limited to the categories of white, black, and "other minorities". The Uniform Selection Guidelines require adverse impact calculations to be made separately for Hispanics, American Indians, and Asians. Therefore, the GAO may not make a finding as to the adverse impact or lack thereof of the PACE and JFA examinations on each of these groups. The GAO conclusions on "other minorities" are thus misleading and should be changed to clarify that no impact data was available. It is, however, highly likely, given other statistics available to the government, that the PACE and JFA examinations do, in fact, have an adverse impact on Hispanics. See, for example, the background material used by Congressman Garcia to support his recent amendment to the 1978 Civil Service Reform Act which established a federal minority recruitment program. Again, the relevant "adverse impact" information still is not being collected, as GAO correctly points out.
- D. The GAO finds that the documentation supporting the validity of the JFA examination is insufficient to meet current standards for validation. Combining this finding with a finding of adverse impact on blacks of the JFA examination, GAO could conclude that the use of the JFA should be discontinued as required by Section 3A of the Uniform Selection Guidelines (restated in IA and IIA of this draft).
- E. GAO finds that the use of scores on the PACE of 90 or above for certification purposes has a greater disproportionate effect on blacks than does use of the scores on a pass/fail basis. GAO could conclude that Section 5G requires evidence sufficient to support this method of use, and specifically, that the technical requirements of Sections 14B(5) and (6) must be met. [See GAO note 2, p. 124.]
- F. GAO finds that no job analyses have been done on 91 of the occupations for which the PACE is currently being used. The job analysis is critical for a construct validity strategy since without the job analysis there is no way to demonstrate the relationship between the test and the job. Since the CSC failed to demonstrate any relationship between scores on PACE and job performances in 91 occupations, GAO could conclude that the use of the PACE for these occupations should be discontinued as required by Section 3A of the Uniform Selection Guidelines and that alternate steps to comply with the Guidelines should be taken. Further, we question the advisability of GAO's recommending continued use of the PACE examination, for other than experimental purposes, given the almost exclusionary impact which the current use of the PACE has on blacks.

IV. GAO recommendations, based on the content of the draft report:

In light of above, we have the following comments about recommendations contained in the draft report:

- A. GAO finds that for 24 of the occupations for which the PACE is used, job analyses have been performed but no criterion-related studies have been conducted. However, we note that the differences between these 24 occupations are great enough for CSC to develop separate Classification Standards and Qualification Standards for each occupation. The Classification Standards detail the job duties and the Qualification Standards detail the knowledges, skills, and abilities which are minimally required by each occupation and at each grade level within each occupation. Given these substantial differences between occupations, strong evidence is needed to demonstrate that a measurement of only five abilities, among the many requisite knowledges, skills, and abilities for each occupation, does in fact predict superior performance on the job. Job analyses, without criterion-related studies, may not be sufficient to demonstrate this relationship. We also question whether the job analyses provide enough evidence of validity to meet the interim use requirements of Section 5J, especially given concerns relating to the three criterion-related studies which are offered as support for the construct validity of the PACE. In addition, the requirement that the method of use of the PACE also be supported by validity evidence intensifies the problem of relying solely on job analyses for demonstrating probable job performance. Given these issues which GAO has yet to address, we question the advisability of GAO's recommending continued use of the PACE examination for these occupations, for other than experimental purposes, given the almost total exclusionary impact which the current use of the PACE has on blacks. [See GAO note 5, p. 125.]
- B. The GAO finds that for 3 occupations for which the PACE is used, the CSC has conducted criterion-related validity studies to demonstrate the relationship between performance on the test and probable performance on the job. The GAO finds that in at least three respects these criterion-related studies fail to meet essential requirements of the Uniform Selection Guidelines. We point out a possible fourth essential which was not discussed by GAO:
1. GAO finds that standards were not met relating to the need for a representative sample in terms of the race and ethnic groups available in the relevant job market.
  2. GAO finds that the requirement for a study of test fairness was not met.

3. GAO finds that the requirement for the investigation of suitable alternatives which would permit the selection of qualified workers while minimizing or eliminating adverse impact likewise was not met.
4. GAO does not discuss an essential requirement for demonstrating the validity of a selection procedure such as the PACE which is used to select people who are expected to progress to a job at a level higher than that for which they are initially selected. The requirement of Sections 5I and 15 G is that if the test measures a knowledge, skill or ability which would be expected to develop principally from training or experience on the job, then use of the test would not be appropriate. Since the three criterion-related studies on the PACE were concurrent validity studies comparing how journey level persons in the three occupations do on the PACE and on the job, the studies may fail to investigate how much is learned in the years a person works in the occupation before reaching journey level status.

[See GAO note 3, p. 124.]

Based on the above findings in its own draft report, the GAO can only conclude that the three criterion-related studies do not meet the requirements of the Uniform Selection Guidelines. Given these facts, we question the advisability of GAO's recommending continued use of the PACE examination, for other than experimental purposes.

[See GAO note 4, p. 124.]

V. Suggested Additional GAO Recommendations:

- A. The most important single failing of CSC's approach to selection procedures is the failure to adequately seek alternative methods of selecting and an equally significant failure to encourage individual agencies to develop and identify alternate selection procedures which will reduce or eliminate adverse impact while continuing to meet legitimate merit system goals. The General Accounting Office report should be amended to contain specific recommendations stressing this problem and urging the Office of Personnel Management to:
  1. actively attempt to develop alternate selection methods which reduce adverse impact;
  2. actively encourage such activities by individual agencies, and
  3. actively make use of the authority in the Civil Service Reform Act of 1978 to conduct innovative and experimental programs for the purpose of developing such alternatives.

This use of alternative selection procedures is an option under Section 6A of the Uniform Selection Guidelines. Note that Section 6A alternatives which have no adverse impact need not be validated in accord with the Guidelines. They would, however, have to meet merit system requirements, including that of "rational validity". For a discussion of this standard see Washington v. Davis. 426 US 229 (1976) in which the Civil Service Commission standards were approved by the Supreme Court in a non-Title VII context with the observation that there was a "rational basis" for the practice challenged in that case. The court went on to suggest that "Under Title VII, Congress provided that when hiring and promotion practices disqualifying substantially disproportionate numbers of blacks are challenged...it is an insufficient response to demonstrate some rational basis for the challenge practices. It is necessary, in addition, that they be 'validated' in terms of job performance." Thus, since the PACE exam has an adverse impact, it must be validated in accord with the Uniform Selection Guidelines. On the other hand, an alternative adopted by the Office of Personnel Management to eliminate adverse impact would (precisely because it has no adverse impact) be supportable under Merit System Standards if there was "some rational basis" for the selection practice adopted to eliminate adverse impact, even if the demonstration of such "rational" basis did not comply in all technical aspects with the validation standards of the Selection Guidelines.

- B. Since the use of the PACE and the JFA examinations have an adverse impact on blacks, and their use has not been validated by CSC nor have the requirements of Section 5J for interim use been met we urge that GAO strongly recommend that the present method of use of the PACE and JFA examinations be discontinued and/or that alternative selection procedures be used which are job related (See discussion in V A above) but which eliminate adverse impact on blacks and other groups. [See GAO note 5, p. 125.]
- C. We note that these recommendations would impact on less than 13% of the applications for federal employment, that is on only 190,000 (135,000 PACE and 55,000 JFA applications) of the more than one and one-half million applications received annually.

VI. Miscellaneous Comments:

In addition to the above general comments a number of smaller items warrant brief comment. They are listed in numbered paragraphs below corresponding to the page number on which the original material appeared in the draft:

[2] #3. The draft report states that:

"Inherent in a merit system is the need to rate and rank individual applicants on the basis of their fitness for the job."

It should be pointed out that the statutory basis of the Federal Merit System (5 U.S.C. 3304) deals with the issue of ranking and relative abilities where it states that competitive examinations must be:

"...practical in character and as far as possible relate to matters that fairly test the relative capacity and fitness of the applicants for the appointment sought." (Emphasis added.)

In determining the extent to which selection on relative ability is "possible" as mandated by this requirement of the merit system one must turn to the technical capabilities of the testing profession in developing selection procedures which are valid for ranking purposes. This is consistent with the requirements of the Uniform Selection Guidelines that it is not sufficient to only validate an instrument as being predictive of job performance but that the particular "use" of the instrument must also be justified. In fact, the first court case arising under these Guidelines Allen v. City of Mobile 18 FEP Cases 217 (1978) specifically held illegal under the Selection Guidelines a validation study which demonstrated that the test was predictive of job performance but which failed to specifically study the use of the test as a ranking instrument.

Thus, the Federal "rule of three" can have meaning only where there is a selection instrument which actually indicates that in terms of job related criteria the "three" are, in fact, "better" in terms of probable job success. Any application of the statutory requirement to select people on relative qualifications "as far as possible" must take into consideration the extent to which it is possible to develop selection devices that can rank in that fashion. [See GAO note 2, p. 124.]

[6] #10. The second sentence on page 10 states that:

"The Uniform Guidelines are first and foremost concerned with assuring non-discrimination; they are not primarily concerned with assuring that qualified persons are hired."

The first part of this sentence is accurate: These are as much guidelines on "discrimination" as they are on "validation" or "testing". In fact, Section 3 begins the major substantive part of the Guidelines by defining discrimination. However, the suggestion that the agencies which issued these Guidelines were not concerned with qualifications is incorrect. Just the opposite is true. It is the goal of these Guidelines to assure that qualified individuals are not excluded from employment. As the Supreme Court pointed out in the Griggs case, it is often inadequate to use "broad and general testing devices... as fixed measures of capability." The Chief Justice suggested that history was filled with examples of individuals who "rendered highly effective performance" without the traditional measures such as tests, certificates, diplomas, or degrees. Rather than disparaging qualifications, what these Guidelines are designed to do is to encourage the identification of individuals who can render the highly effective performance the Chief Justice mentioned but who may not be identified by the selection procedures currently in use. The second part of the quoted sentence should be deleted or amended.

[8]#13. Four lines up from the bottom the report adds the phrase "other racial minorities" in handwriting. Under Section 4b of the Guidelines it is inappropriate to aggregate minorities in this fashion. [See GAO note 1, p. 124.]

[20]#31. The last three lines on page 31 suggests that "... we do not know whether Civil Service employment procedures resulted in adverse impact as administratively defined by the new Uniform Guidelines." As pointed out in the earlier part of this letter, this is incorrect. The Guidelines suggest that even where there is adverse impact there may be times where, as in a matter of administrative and prosecutorial discretion, the agencies will not take action, but the Guidelines do not suggest that in those circumstances adverse impact does not exist.

- [22] #34. The first three lines in the first full paragraph suggest that the existence of adverse impact has not yet been determined. See previous comment.
- [25] #36. The recommendation contained in the next to last paragraph should recommend the maintenance of records which will disclose impact on employment opportunities of "specified minorities and sex groups". (See comment re page 13, above).
- [26] #37. The last line and one half on the page stating "we found that a substantial amount of evidence has been gathered to support the validity of the PACE..." suggests the report believes that the PACE exam has been validated when a careful reading of the paragraph and of the rest of the report indicates this not to be the case. This should be revised to read something like:
- "GAO found that the Civil Service Commission had gathered a substantial amount of evidence in an attempt to demonstrate the validity of the PACE..."
- [26] #38. After the words "and job relatedness" in the third line the phrase "or preferably, take steps to eliminate adverse impact, or develop alternatives which eliminate adverse impact" should be added.
- [29] #42. In line seven there is the suggestion that criterion-related validity studies are not technically feasible. We refer GAO to Section 16U defining technical feasibility and suggest that the problem raised on page 42 has more to do with the issue of administrative and financial capability rather than the issue of technical feasibility which deals with the size of sample, range of scores, availability of appropriate criteria and other matters focusing on the technical aspects of the validity study rather than its size and cost. Actually, the very size which makes it more costly increases its technical feasibility.
- [35] #49. There are several problems with the discussion of certain measures of job performance for research participants entering Federal employment through written test (PACE/FSEE) or through "an alternative selection procedure" (emphasis added).
- First, it suggests that an adequate search for alternatives to comply with the requirements of Section 16X and 3B has been conducted without discussing the nature of the search. We question its adequacy.

- It does not identify the nature of the single "Alternative Selection Procedure" discussed.
- It does not give us information concerning possible qualified individuals rejected by the written test and what they might have done on the Alternative Selection Procedure and on the measures of job performance.

[See GAO note 7, p. 125.]

#53 [40] The suggestion in the first line of the first full paragraph that "construct validity has been most widely used" is factually incorrect. It is widely used by the Federal Government in support of the PACE exam. It is almost never used by private employers and very seldom by other employers at the State and Local government level. Thus, that sentence should read:

"For these reasons CSC, has used construct validation substantially and has on some occasion funded the use of the construct validity strategy by State and local governments. However, construct validity strategy is almost never used in the private sector and evidence of its use by State and local governments, except where financed by the Civil Service Commission, is limited." [See GAO note 6, p. 125.]

We appreciate the opportunity to comment on your draft report and have confidence that you will give our recommendations your full consideration.

Sincerely,



Eleanor Holmes Norton  
Chair

GAO notes: To the extent time permitted, we considered and discussed EEOC's comments in this final report. However, because of the timeframe involved for publishing the final report, we did not discuss or respond to each comment in the body of the report. Where we believed additional commentary was needed but not specifically provided in the report, we provided the following GAO notes.

1. We recognize that the Guidelines require that records be kept to show whether adverse impact is present for each specific minority group. However, such records had not been kept. Therefore, we used information available at the Social Security Administration which was categorized as "white," "Negro," and "other." Information on the composition of the "other" category is not available. (See p. 8 of the report.)
2. The technical requirement referred to concerns criteria for determining statistical significance and the need for stronger correlations with a greater number of job elements when rankings are used and adverse impact exists. The PACE criterion-related studies showed such a correlation between the test and multiple measures of job performance. Thus, we believe that the evidence shows that the requirements of section 5G on method of use have been met.
3. The types of abilities measured by PACE--reasoning, judgment, numeric, and verbal--would not likely be developed primarily from job training or experience.
4. Based on the evidence examined, we do not believe a conclusion that the use of PACE should be discontinued is warranted. While we noted shortcomings in OPM's research, we also noted a substantial amount of carefully performed research work. As evidenced by our recommendations, we believe more research work should be done, but we do not believe the work already performed should be disregarded.

5. The Uniform Guidelines, section 5J, state that a selection procedure which is not at the moment fully supported by the required validity evidence may continue to be used provided that there is available substantial evidence of validity and that a study is in progress designed to produce the additional evidence required by the Uniform Guidelines. Based on our evaluation, we believe that PACE meets the interim use requirements of the Guidelines.
6. In our opinion, the report statement is correct. When the construct validity method is used, it is most often used in mass selection situations.
7. Deleted comment relates to a matter in the draft report but deleted from the final report.
8. Page numbers in brackets refer to pages in this report.

UNITED STATES COMMISSION ON CIVIL RIGHTS  
WASHINGTON, D. C. 20425



STAFF DIRECTOR

APR 11 1979

Mr. Gregory Ahart  
Director  
Human Resources Division  
General Accounting Office  
Washington, D. C. 20506

Dear Mr. Ahart:

Chairman Arthur S. Flemming has asked me to thank you for providing us with copies of "Federal Employment Examining and Selection Procedures: Do They Achieve Equal Opportunity and Merit Principle Goals?" We are pleased to give you our comments on this innovative study.

Naturally we were disturbed by the study's finding that only 0.2 percent of black applicants achieved scores high enough for a realistic chance of being considered for employment, while 16 percent of white applicants obtained such scores. In our 1977 Enforcement Effort report, To Eliminate Employment Discrimination: A Sequel, we concluded from available evidence that the PACE was a major barrier to Federal Employment for minorities. The General Accounting Office (GAO) results revealed that the PACE restricts employment opportunities for blacks to a far greater extent than we had suspected.

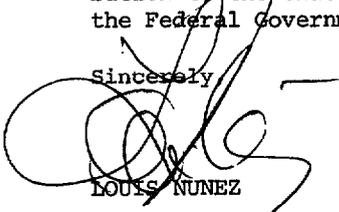
Thus, regardless of the validity of the PACE, it is clear that few minorities are able to enter Federal employment through the PACE. GAO should urge the Office of Personnel Management (OPM) to consider replacing the PACE with a less discriminatory alternative. GAO has demonstrated that some unassembled examinations do not exclude minorities to the same degree as the PACE, and we suggest that such examinations are an alternative which should be thoroughly explored.

If OPM is to continue using the PACE, however, it is imperative for the Government to determine as quickly as possible whether its use comports with the validation requirements of the Uniform Guidelines. GAO recommends that the Director, OPM, reevaluate the PACE validation strategy. However, inasmuch as OPM believes that the PACE has been adequately validated and that OPM's validation methodology is sufficient, we are convinced that it is necessary for an independent and impartial review of the validity of the PACE. Therefore, we urge GAO to recommend that the Equal Employment Opportunity Commission (EEOC) undertake such a review, consistent with EEOC's responsibilities under the President's Reorganization Plan No. 1 of 1978.

Finally, we wholeheartedly concur with the GAO draft recommendation that the Director, Office of Personnel Management (OPM), "immediately act to comply with the Uniform Guidelines on Employee Selection Procedures requirement for collecting and maintaining records which will disclose the impact that tests and other selection procedures have on the employment opportunities of minority group members." If such a system were in place, the Federal Government would know the impact of its selection procedures and would be in a position to eliminate any adverse impact. If GAO believes that a comprehensive system for collecting and analyzing racial and ethnic data on applicants cannot be fully implemented immediately, GAO should recommend, as an interim step, that OPM obtain racial data from the Social Security Administration files, as was done in this GAO study.

We appreciate being asked to review this report, and we hope that you find our comments to be useful. GAO's research on the PACE makes a real contribution to the existing body of knowledge on equal employment opportunity in the Federal Government.

Sincerely,



LOUIS NUNEZ



## UNITED STATES DEPARTMENT OF JUSTICE

WASHINGTON, D.C. 20530

Address Reply to the  
Division Indicated  
and Refer to Initials and Number

APR 23 1979

Mr. Allen R. Voss  
Director  
General Government Division  
United States General Accounting Office  
Washington, D.C. 20548

Dear Mr. Voss:

This letter is in response to your request for comments on the draft report entitled "Federal Employment Examining and Selecting Procedures: Do They Achieve Equal Opportunity and Merit Principle Goals?"

The Department of Justice defers making any comments on the draft report at this time because of the pendency of two cases in the district courts, both of which challenge the lawfulness of the use of the Professional and Administrative Career Examination. The Civil Division is representing the Office of Personnel Management and there is some possibility the cases can be settled. Any substantive comments on the report would be inconsistent with the Department's role as lawyer and might prejudice the possibilities for settlement.

We appreciate the opportunity given us to comment on the draft report. Should you desire any additional information, please feel free to contact us.

Sincerely,

A handwritten signature in cursive script that reads "Kevin D. Rooney".

Kevin D. Rooney  
Assistant Attorney General  
for Administration



(964095)

Single copies of GAO reports are available free of charge. Requests (except by Members of Congress) for additional quantities should be accompanied by payment of \$1.00 per copy.

Requests for single copies (without charge) should be sent to:

U.S. General Accounting Office  
Distribution Section, Room 1518  
441 G Street, NW.  
Washington, DC 20548

Requests for multiple copies should be sent with checks or money orders to:

U.S. General Accounting Office  
Distribution Section  
P.O. Box 1020  
Washington, DC 20013

Checks or money orders should be made payable to the U.S. General Accounting Office. NOTE: Stamps or Superintendent of Documents coupons will not be accepted.

**PLEASE DO NOT SEND CASH**

To expedite filling your order, use the report number and date in the lower right corner of the front cover.

GAO reports are now available on microfiche. If such copies will meet your needs, be sure to specify that you want microfiche copies.

**AN EQUAL OPPORTUNITY EMPLOYER**

**UNITED STATES  
GENERAL ACCOUNTING OFFICE  
WASHINGTON, D.C. 20548**

**OFFICIAL BUSINESS  
PENALTY FOR PRIVATE USE, \$300**

**POSTAGE AND FEES PAID  
U. S. GENERAL ACCOUNTING OFFICE**



**SPECIAL FOURTH CLASS RATE  
BOOK**