

109388
~~101166~~
UNITED STATES GENERAL ACCOUNTING OFFICE
WASHINGTON, D.C. 20548

FOR RELEASE ON DELIVERY
EXPECTED AT 9:30 a.m. EDT
Tuesday, May 15, 1979

STATEMENT OF

CLIFFORD I. GOULD, DEPUTY DIRECTOR
FEDERAL PERSONNEL AND COMPENSATION DIVISION

BEFORE THE

SUBCOMMITTEE ON CIVIL SERVICE
HOUSE COMMITTEE ON POST OFFICE AND CIVIL SERVICE

ON

THE IMPACT AND VALIDITY OF PACE: A FEDERAL EMPLOYMENT EXAMINATION

Madam Chairwoman and Members of the Subcommittee.

We appreciate the opportunity to discuss our report on Federal employment examinations and the question of whether the examinations achieve equal opportunity and merit principle goals. Our report, released today, focuses on a very perplexing problem confronting those responsible for ensuring that the Federal work force is competent, productive, and to the extent possible, reflective of all segments of our Nation's society.

The goal of Federal employment is to hire qualified individuals based on merit principles giving everyone an equal opportunity to compete for a job. To achieve this goal the Civil Service Commission designed various examining



109388

005307

and selecting procedures to help assure that the most competent and productive people are employed. This responsibility now belongs to the Office of Personnel Management (OPM).

The problem which has perplexed managers in Government and other concerned observers is that black job applicants are screened out of the competition at a much higher rate than white applicants as a result of their relative performance on two major written tests used by OPM to develop hiring registers. Our study of PACE (Professional and Administrative Career Examination) for example, showed that over 99 percent of all blacks who took the test during the period were effectively screened out of competition because they did not score 90 or better on the test, the score usually needed for Federal job certification. By comparison, only 84 percent of white applicants were screened out.

What makes this situation perplexing is that PACE is the most thoroughly researched test in the history of Federal civil service examining. According to OPM the test has been fully validated and its research shows a clear relationship between performance on the test and subsequent performance on the job. Therefore, OPM says that not only does the test result in hiring individuals who perform better on the job, but its use enhances overall Government productivity.

Critics of PACE have focused on the validation strategy followed by OPM. They point out that in some areas the research work on the test does not strictly conform to the Federal Government's Uniform Guidelines on Employee Selection Procedures--although the Guidelines were not in affect when the test was developed. Our study revealed shortcomings in OPM's validation work, and we made recommendations aimed at overcoming those shortcomings.

Our report includes information on three examinations in addition to PACE--one written and two "unassembled" examinations. Because your hearings this morning are concerned primarily with PACE, my testimony is confined to our study of that examination.

PACE is used to hire new employees for entry level jobs in 118 different occupations. According to OPM, the most common feature of PACE occupations is their information burden--that is, all of the occupations require individuals who can read, understand and communicate a large body of written laws and regulations or a large body of knowledge related to a specific profession or specialist area.

Using procedures I will describe later, OPM determined that five abilities were important to successful performance in these occupations and could be included in a written test. These were:

- Verbal comprehension: The ability to understand and interpret complex written material, and to use language where precise use of words and concepts is necessary for effective communications.
- Judgment: The ability to make correct decisions in the absence of complete information and to solve problems by inferring missing facts.
- Deductive reasoning: The ability to discover underlying relationships among specific data where solving problems involves forming and testing hypotheses.
- Inductive reasoning: The ability to reason to general principles from specific situations.
- Quantitative reasoning: The ability to perform arithmetic operations and solve quantitative problems where the proper approach is not specified.

Individuals who pass the test and receive a rating of 70 or more are listed on a PACE job register according to their final rating--test score plus points for Veterans' preference or for being an outstanding scholar, when applicable. When entry level jobs become available in one of the 118 occupations, agencies request OPM to send them a list of names from the PACE register. In response, OPM sends the agency the three names with the highest ratings. Agencies generally must select one of the three individuals on the list, and if one of the individuals on the list is an

eligible Veteran, the Veteran must be selected. The process is somewhat more complex than what I have just described, but this should help explain what is involved for an individual who has not had a Government job in obtaining an entry level PACE position.

Before I discuss the results of our study, I would like to put PACE in perspective in terms of its importance in obtaining a Federal job. During fiscal year 1978, OPM processed 1.6 million job applications for all types of competitive civil service jobs, and 152,771 people were hired. These numbers include all competitive jobs, both wage grade and general schedule. During that same year 135,402 individuals took the PACE, 76,907 passed with a score of 70 or above, and 7,587 were hired from PACE registers. While PACE accounted for only about 5 percent of the individuals hired in 1978, the test is important because it is the entry route into Federal service for more college graduates than any other single method and because the type of jobs covered by the test often leads to a higher level career position.

Our study and report focused primarily on the first part of the selection process; that is, the PACE examination. We wanted to determine: Do white and minority applicants score at about the same level on the examination? If not, does PACE have adverse impact on minorities? And,

has the test been properly validated? We did not specifically review the process of sending names to agencies, or the process agencies go through in selecting employees from the lists of names they receive.

DO WHITES AND MINORITIES

SCORE AT THE SAME LEVEL?

Because OPM had not maintained information on the race or ethnic group membership of job applicants, the information needed to answer that question was not available from their files. To get information, we matched OPM's data on PACE applicants with the Social Security Administration's master file. The master file did not contain all the information we would liked to have had, but it did tell us whether the PACE applicant had identified himself or herself on their social security card application as "white," "Negro," or "other" race. The Social Security Administration's file did not tell us the composition of the "other" racial group category, nor did it provide information on different ethnic groups. Therefore, we could not determine how other specific minority groups such as Hispanics or American Indians performed on PACE. We did, however, obtain information on black applicants.

We analyzed performance on PACE using three different approaches: (1) augmented scores--the test score plus Veterans' preference points and the outstanding scholar

computation; (2) raw test results; and (3) a separate analysis of applicants who claimed outstanding scholar status. The results were the same for each approach. That is, the test screened out black applicants at a much higher rate than either white applicants or applicants who had identified themselves as coming from "other" racial groups.

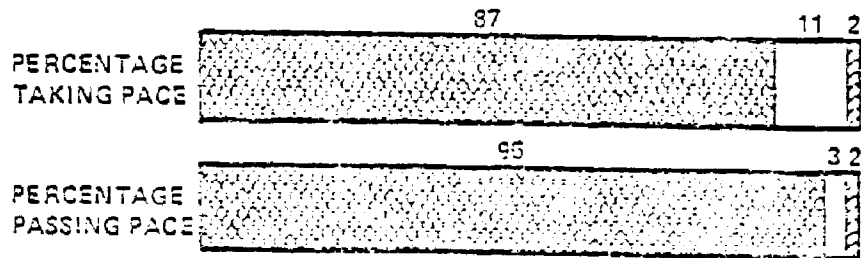
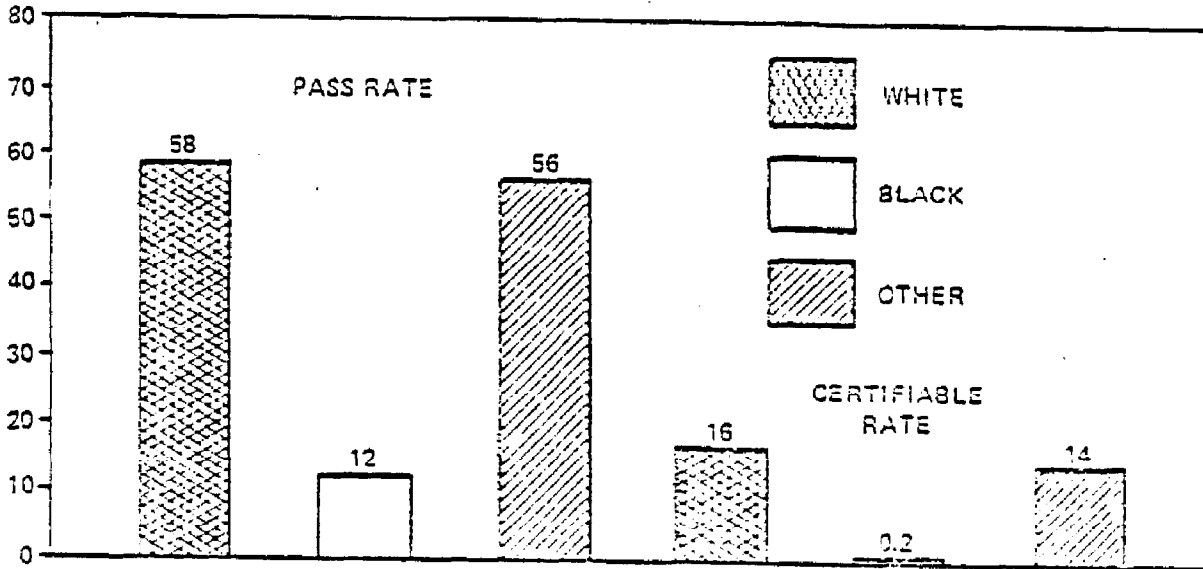
Our analysis of augmented test scores--the rating used for purposes of ranking individuals on job registers--showed that:

--58 percent of white applicants passed; 16 percent with a score of 90 or above,

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

--12 percent of the black applicants passed; less than 1 percent with a score of 90 or above.

This is shown graphically below.



OPM officials told us that, depending upon job market conditions, a score of at least 90, and sometimes higher, is needed for an individual to have a realistic chance of having his or her name sent to an agency for employment consideration.

One aspect we were particularly interested in was how well black and white applicants who claimed outstanding scholar status performed on the test in relationship to one another. Outstanding scholar status may be claimed by individuals who graduate with a 3.5 grade point average--on

a 4 point scale--or in the upper 10 percent of their class. We wanted to know if people with roughly equivalent educational standings performed equally well on the test. We found that they did not. Our analysis showed that 53 percent of white outstanding scholars, but less than 1 percent of black outstanding scholars scored 90 or above on the test.

DOES PACE HAVE ADVERSE IMPACT?

In view of the results of our analysis of PACE scores, the answer to that question would seem obvious. The answer is important because it determines whether the technical standards of the Uniform Guidelines should apply to the validation of the test. However, the answer is not as clear cut as it may seem. The Guidelines define adverse impact as a substantially different rate of selection which works to the disadvantages of members of a race, sex, or ethnic group, but it states that for purposes of determining when to take enforcement action, adverse impact will usually be determined based upon the total selection process for a job rather than upon the impact of one component of the process. This is known as the "bottom line" concept. The Guidelines, however, do not rule out enforcement action if one component has adverse impact.

A disproportionately large percentage of black applicants are screened out as a result of their performance on

the PACE written test. However, PACE is not the only "port-of-entry" into an occupation it covers. OPM has estimated that only about a third of the people currently holding jobs in PACE occupations entered as a result of taking PACE or its predecessor examination, the Federal Service Entrance Examination. We don't know how reliable OPM's estimate is, but we do know that there are a number of different "ports-of-entry" into a PACE occupation which do not require taking the examination. These "ports-of-entry" include promotions, reassignments from related occupations within an agency or between agencies, reinstatement of former Federal employees, Upward Mobility Programs, Cooperative Education Programs, and the Presidential Management Intern Program.

OPM has also estimated that minority group members comprise about 17 percent of incumbents in GS grades 5 to 11 in 16 of the most populous PACE occupations. OPM said that this 17 percent figure is higher than relevant labor market statistics and also higher than the percentage of minorities in comparable occupations in the private sector. Therefore, in OPM's opinion, there is no "bottom line" adverse impact on minorities in PACE occupations.

While we agree with OPM that the incumbent population is relevant in determining when to take enforcement action, the Guidelines state that adverse impact is computed based

on the rate of selection from the applicant population, not on the basis of the number of minorities currently employed. If, as OPM maintains, 17 percent of the incumbents in PACE occupations are minorities--blacks, Hispanics, American Indians, Asian Americans, and others--but, the test screens out such a high proportion of blacks, there is a question as to whether two standards prevail--one for whites (PACE) and another for minorities through some other method.

The reason that OPM cannot determine if adverse impact exists in PACE occupations using either the "bottom line" concept, or based on an individual component of the selection process, is because it has not maintained records to show the race or ethnic group membership of applicants. We have recommended that OPM collect this necessary data. A requirement to collect this type of data has been in effect since 1976, and it is a requirement of the Uniform Guidelines.

OPM has been working on a data collection system since 1976, but they have had problems in determining how to obtain reliable race and ethnic data from applicants. Also, according to OPM, they have had problems getting funds allocated for this project and the Office of Management and Budget (OMB) has not approved the use of the form needed

to collect the data. OPM told us that they have now resolved the problem of how to collect reliable data on applicants who take written tests. We believe the time has come for OPM to give higher priority to funding this project, and for OPM and OMB to resolve their differences regarding the use of the form needed to collect the data.

The Equal Employment Opportunity Commission (EEOC), the agency responsible for enforcing the Uniform Guidelines, reviewed the facts presented in our report and concluded that even if OPM chooses to defend PACE on the grounds that the "bottom line" of the total selection process is good, it is precluded from doing so at this time because of its failure to collect the relevant adverse impact information.

HAS PACE BEEN PROPERLY VALIDATED?

After we had determined that PACE screens out a high proportion of black job applicants, we examined whether PACE was a valid test for the 118 occupations for which it is used. We wanted to know whether a clear relationship between test scores and performance on the job had been shown. We found that OPM has spent a great deal of time and money developing PACE and performing research on its validity. OPM maintains that PACE is a fully validated test for the 118 occupations. This a highly technical area and I will not take the time to describe all the details of what OPM did to validate the test, but will try to briefly describe their work.

There are three methods of test validation that have been generally accepted by the psychological profession and recognized by the Uniform Guidelines. These are content validity, criterion-related validity, and construct validity. Construct validity--the method used by OPM for developing PACE--is demonstrated by showing that the selection procedure measures the degree to which the job candidates have the characteristics--constructs--which have been determined to be important for successful job performance.

The link between a selection instrument and a job or group of jobs using construct validation is essentially a two step process to see if the test measures what it is supposed to measure and to see if the abilities measured are important to successful job performance. The first step is to analyze the jobs to determine what duties are performed on the job and what knowledge, skills, and abilities are important to successful performance of those duties. The second step is to develop a test which measures the important knowledge, skills, and abilities, and to show that the test in fact measures the abilities determined to be important to job performance. The process of showing that the test measures the important abilities consists of performing a series of criterion-related validity studies and making judgments of validity based on a preponderance of research evidence developed on tests like the one being researched--in this case, PACE.

While my description of this two step process may sound relatively simple it is in fact an extremely complex and arduous process and must be carried out with great care and scientific rigor.

The first step in the PACE validation process was to identify the constructs important in job performance and to determine how they would be measured. In order to do this, 27 occupations were selected for intensive study. Approximately 70 percent of the annual PACE appointments in prior years had come from these 27 occupations. Senior-level and supervisory employees in the occupations developed lists of duties and then rated their importance and the amount of time spent performing them. Over 1,200 "subject-matter experts" from 26 Federal agencies provided these ratings. They also rated a set of knowledges, skills, abilities, and other characteristics in terms of their importance for successful job performance. These ratings provided one basis for identifying constructs to be measured by the written test.

The other basis for identifying constructs was the psychometric literature on aptitude and achievement measurement. The modern literature on this subject began in the 1930's and continued through World War II when the Army Air Corps developed and analyzed hundreds of tests for its Aircrew Classification Battery. After the war,

other tests were developed to measure mental ability factors including the General Aptitude Test Battery used by the United States Employment Service.

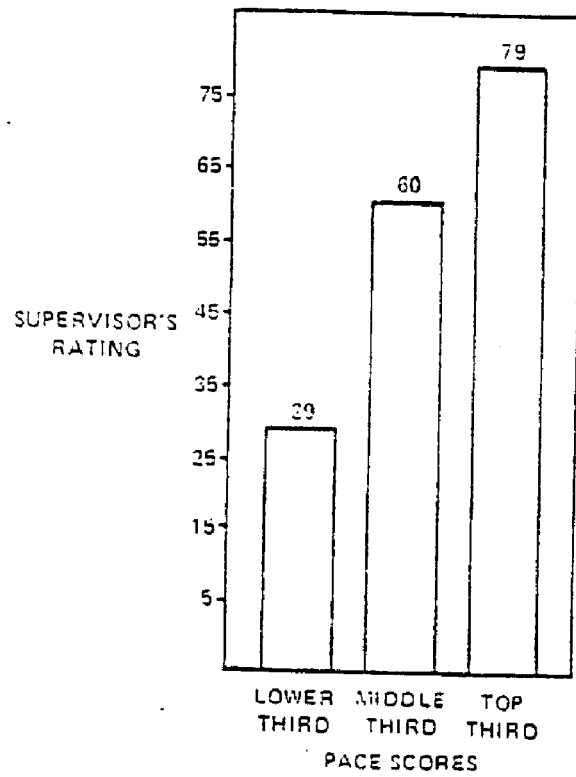
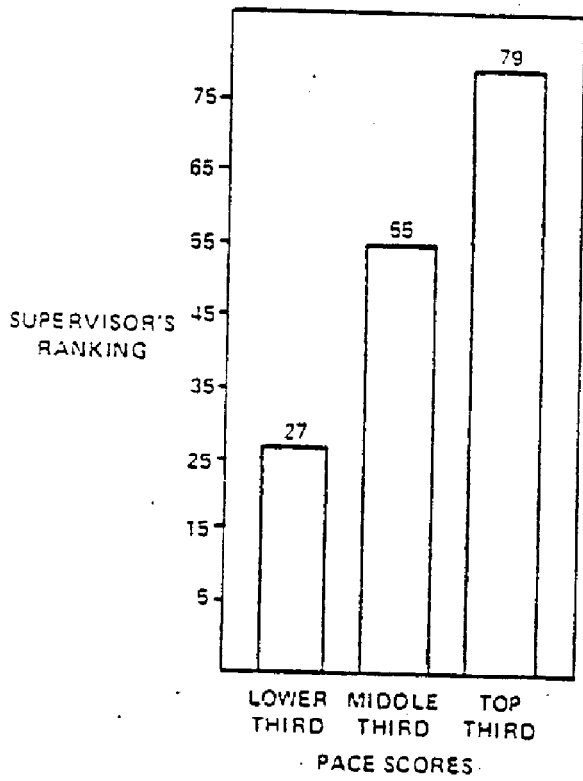
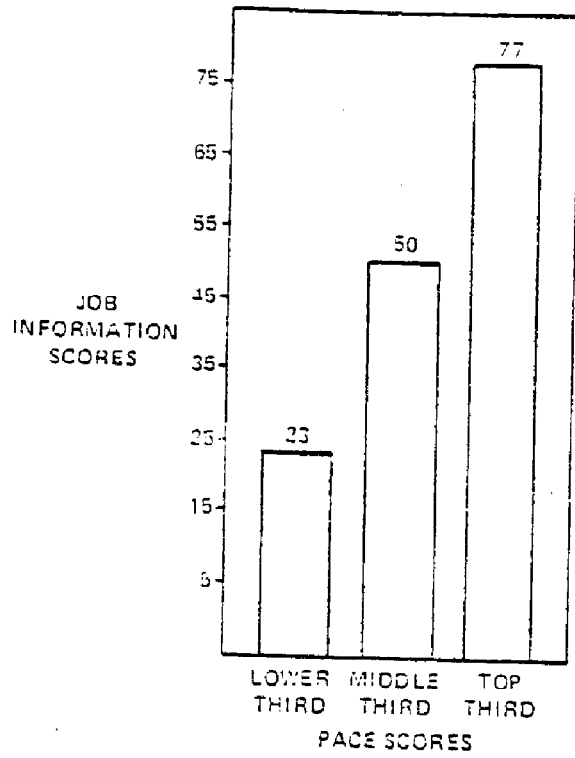
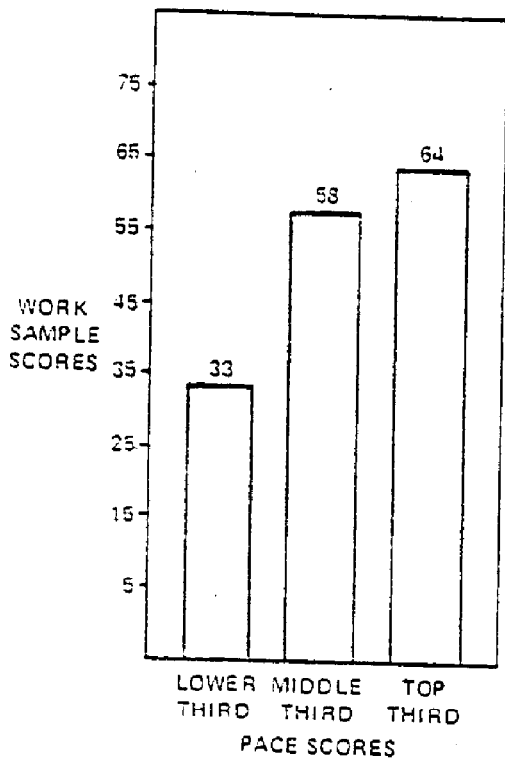
A comprehensive review in the early 1950's identified hundreds of tests whose construct validity had been explored in diverse settings. OPM researchers matched the knowledges, skills, and abilities rated by the subject matter experts with the constructs identified in this comprehensive review. They then wrote test questions similar to those contained in earlier tests which measured those constructs.

As soon as the PACE written test was developed, OPM started follow-up research. One aspect of that research involved using criterion-related validity studies to determine the extent to which the constructs were related to job-performance in very populous occupations.

The first of the criterion-related validity studies was conducted on social security claims examiners. The written portion of PACE was administered to a sample of 250 examiners and measures of their work effectiveness were collected. These measures were specially designed for the study and covered many facets of an examiner's work. They consisted of a job information test, a work sample consisting of a standardized claim to be adjudicated by the examiner, a rating form completed by a supervisor, and a

supervisory ranking form. A 528-statement task inventory was used to obtain a description of the work to provide a basis for developing the measures of job performance.

OPM's analysis of the relationships between test scores and job performance measures showed that the test was valid--that is, the correlation between scores on the test and measures of job performance was statistically significant. The graph below gives some indication of this relationship. It shows, for example, that those who scored in the bottom third on PACE attained work sample scores which were at least 26 points lower than those who scored in the top third on PACE.



Studies of the criterion-related validity of the written portion of PACE have been completed in the internal revenue officer and customs inspector occupations. The design of these studies paralleled the design of the claims examiner study--that is, a detailed job analysis was done and measures of job performance were developed; geographically dispersed samples were selected; and measures of job performance were compared with performance on the examination. In almost all of the comparisons between test scores and job performance measures OPM found the test to be valid.

Professional standards hold that construct validity is established by a preponderance of evidence from a variety of sources. In view of this, OPM has published a paper which summarizes research done in the last 30 years on criterion-related validity of constructs measured by the PACE. OPM's summary was limited to jobs similar or identical to those for which selections are made through the PACE. It included 126 validity studies which presented a wide variety of job situations. Almost without exception each study reported a statistically significant relationship between the measure of the construct and the measure of job performance.

As I have indicated, OPM has an extensive amount of documentation to support its contention that the PACE is valid. We evaluated that documentation in light of (1) the

impact the test has on blacks, (2) the various criticisms that have been voiced regarding the validation strategy, (3) professionally acceptable methods of the psychological profession for demonstrating whether a selection procedure validly predicts job performance, and (4) the Uniform Guidelines on Employee Selection Procedures issued by the Federal Government. We concluded that, while OPM's validation work generally conforms to professional standards, there were some shortcomings and two standards described as "essential" were not followed. Also, in some areas the validation work does not now conform with the Uniform Guidelines. We believe that since OPM signed the Guidelines which almost all other employers--public and private--must follow, it cannot afford to do less than that which is required of others.

One of the criticisms voiced about PACE is that it is now used for 118 occupations but detailed job analyses were performed for only 27 occupations. Consequently, the critics say that OPM has not adequately demonstrated that the remaining 91 occupations covered by the test have critical work behaviors common to the 27 initially studied.

In commenting on our report OPM said that this criticism is unfounded and that they did in fact perform job analyses of the 118 occupations.

Because of OPM's statement, we looked further into what had been done in the area of job analysis for these remaining 91 occupations. We found that at the time the original research and development work was performed, OPM had planned to perform a comprehensive analysis of the remaining occupations not included in the initial research. This analysis, while it was not to have been as intensive and detailed as that performed on the 27 occupation, was to have followed the same general study plan. It was to have involved the use of subject matter experts who would rate the importance of duties, time spent on each duty, and the importance of abilities required for performing the duties. Because this analysis was planned for the future, during 1974, as an interim measure, an occupational specialist familiar with PACE and the qualification and classification standards of the occupations, spent a few days evaluating the remaining occupations in the context of the five abilities tested by PACE. This evaluation resulted in some occupations being dropped from PACE coverage and some being added. However, the comprehensive analysis originally planned for the remaining occupations was never performed.

The Uniform Guidelines clearly require a thorough and fully documented job analysis when a test like PACE is used. In our opinion, the interim evaluation does not satisfy

these requirements. Therefore, we have recommended that, if PACE continues to be used for these 91 occupations, the originally planned comprehensive job analysis should be completed.

OPM has also been criticized for using the construct method to validate the test rather than using some other method. This method is the most difficult of the three accepted validation strategies, but this method also has the highest payoff for a large employer with many occupations requiring similar skills or job behaviors, and where there are far more applicants than jobs. In our opinion, construct validation was an appropriate method for OPM to use in validating PACE. It is an appropriate method to use with a test which measures abilities needed for success in a large number of jobs where no specific prior job knowledge is needed or expected, without performing criterion-related studies for each job. Also, a single test which measures abilities needed in a large number of jobs enhances the opportunities of Federal job seekers. PACE, for example, allows an individual to take one test to qualify for 118 occupations rather than taking 118 tests for a job in an occupation the individual may not even know exists.

OPM has the professional resources to conduct construct validity studies and we believe it exercised due professional

care in carrying out the study. However, in performing its supporting criterion-related studies two standards described by the American Psychological Association and the Uniform Guidelines as "essential" were not followed. These standards concern the need to collect race and ethnic data on the research sample and the need to investigate test bias or test fairness. The APA standards and the Uniform Guidelines also say that, to the extent feasible, the research sample should be representative of the relevant labor market. This also was not done. In our opinion, since the test screened out such a high percentage of blacks as compared to whites, all of the standards should be met. We have recommended that OPM follow each Uniform Guidelines requirement if it continues to use PACE.

Critics of PACE also contend that the design of the criterion-related studies was inadequate. These studies used the concurrent design--using current employees as research participants--rather than the predictive design where job applicants are used. The Division of Industrial-Organizational Psychology of the American Psychological Association has said that predictive studies are preferred from a scientific standpoint. They also said that concurrent studies cannot be expected to answer questions of prediction but that under appropriate circumstances, data obtained in a concurrent study can be used to estimate the predictive validity of a test.

OPM told us that predictive studies were not performed because civil service laws require that applicants be ranked according to their scores and fitness for the job. This viewpoint, of course, presupposes that the test being validated is valid. While we can agree with OPM that a predictive study in the theoretical sense might not be technically feasible because every applicant cannot be hired, we believe it may be possible to conduct a study where individuals who enter a PACE job through some other "port-of-entry" are given the test as part of the research project and their job performance is followed up at a later date. We have recommended that OPM consider using this approach if further research is done on PACE. OPM said that it is planning such a study with the Social Security Administration.

Another criticism voiced about OPM's research work is that it did not adequately investigate for suitable alternatives to PACE which have less adverse impact.

The Uniform Guidelines require that the validation study include an investigation of suitable alternatives, or alternative ways of using the test, which have less adverse impact. "Suitable" is defined in the Guidelines as a selection procedure which is "substantially equally valid." OPM, during their research effort explored alternative ways of using PACE and some alternative selection methods which

would have less adverse impact. However, they told us that they did not find any alternatives that were as valid, cost-effective, and/or as practical as PACE.

We are not convinced that sufficient investigation has been done to find suitable alternatives to PACE, at least for some occupations, or to find suitable alternative ways to use the test which has less adverse impact. There are many people entering PACE jobs without taking the test, and according to OPM's estimates, many of these are minorities. If OPM's estimates are correct, and these alternatives "ports-of-entry" have less adverse impact, it would seem reasonable for OPM to investigate these alternatives to determine whether they are "suitable". Again, assuming that OPM's estimates are correct, it seems inconsistent that one group of individuals--primarily whites--enter through PACE, and another group--minorities--enter through some other method.

We have recommended that OPM design, as part of the recordkeeping system required by the Uniform Guidelines, a procedure to track the job performance of groups of people selected from civil service registers and those selected using alternative "ports-of-entry." If the tracking system shows that one or more of the alternative "ports-of-entry" have less adverse impact on blacks than PACE and result in the selection of employees whose job performance is as good

or better than those selected from registers, we believe OPM should consider expanding the use of these alternatives. If the opposite is found, that is individuals selected by alternative procedure do not perform on the job as well as those selected from registers, we recommended that OPM re-examine the alternative procedure to determine whether it should continue to be used.

In summary, our report focuses attention on a very perplexing problem for which there is no easy solution. The Government is using a test--PACE--that, on the one hand, screens out a disproportionately large percentage of blacks from competition for a Government job. Thus it inhibits achievement of the goal of having a Federal work force reflective of all segments of society. On the other hand, OPM believes that the test has been fully validated; that a clear relationship between test scores and job performance has been shown; and that the use of the test enhances overall productivity of the Federal work force.

This completes my prepared statement. We will try to answer any questions you or other members of the Subcommittee may have.