A general discussion is presented of 2 phases of a study designed to determine the relationship between selection techniques (used by the State of Colorado Civil Service Commission and the Career Service Authority of the city and county of Denver) and on-the-job performance in specified job classifications for personnel classified as Negro, Spanish-surname, or white. Jobs included in the longitudinal analysis were (1) Hospital Attendant at Denver General Hospital, (2) Intermediate Clerk Typist, (3) Clerk Stenographer, Intermediate Clerk Stenographer, and Senior Clerk Stenographer, (4) Clerk Typist, Senior Clerk Typist, and Dictation Machine Operator (all in various State office settings), and (5) Resident Supervisor Trainee at Lookout Mountain School for Boys. Two questions were asked: (1) Is there evidence of unfair discrimination practices in employment selection? and (2) How accurate is the probability of success on the job as assigned to individuals in the applicant population at the time of application? No evidence of unfair discrimination in terms of ethnic group membership was found, but it was noted that it should be possible to improve selection accuracy by refining the process by which the "probability-of-success-on-the-job-if-hired" is determined. A related document is RC 004 240. (DK)
INTRODUCTION

This report contains the findings of a two-year project designed to determine the relationship between selection techniques used by the State of Colorado Civil Service Commission and the Career Service Authority of the City and County of Denver and on-the-job performance in specified job classifications when personnel are classified by (1) Negro, (2) Spanish-surname, and (3) White groups.* The project, conducted in two phases, involved (1) the cross sectional testing of present employees and the concurrent assessment of their job performance, and (2) the longitudinal testing of job applicants with a subsequent follow-up of their on-the-job performance.

The jobs involved in the cross sectional analysis included (1) Hospital Attendant at Denver General Hospital, (2) Resident Supervisor at Lookout Mountain School for Boys, (3) Retardation Technician at the Colorado State Home and Training School at Wheatridge, (4) Food Service Worker, and (5) Psychiatric Technician at Pueblo State Hospital.

The jobs involved in the longitudinal analysis included (1) Hospital Attendant at Denver General Hospital, (2) Intermediate Clerk Typist, (3) Clerk Stenographer, Intermediate Clerk Stenographer and Senior Clerk Stenographer, (4) Clerk Typist, Senior Clerk Typist and Dictation Machine

*When the present project was initiated in 1967, the designations Negro, Spanish-surname and White groups were in widespread use at that time. These designations have been retained throughout the project to assure consistency despite the shift in popular designations for these groups. Synonymous designations for the three groups include Afro-Americans, Hispanos, Spanish or Mexican Americans and Anglos.
Operator (all in various State office settings), and (5) Resident Supervisor Trainee at Lookout Mountain School for Boys.

BACKGROUND

When the Tower Amendment to the Civil Rights Act was passed, the legality of using employment tests was noted, but the responsibility of employers to demonstrate that their selection tests would not result in unfair discrimination associated with race, color, religion, sex or national origin was clearly implied:

"It shall not be...an unlawful employment practice for an employer to give and act upon the results of any professionally developed ability test provided that such test...is not...used to discriminate because of race, color, religion, sex, or national origin."

Civil Rights Act, 1964, Section 703(h)

Fulfillment of the responsibility for demonstrating that his tests are not unfairly discriminatory requires that the employer know (1) how well individuals from various ethnic subgroups of the population can perform the job in question, (2) how well individuals from various ethnic subgroups of the population perform on the tests being used, and (3) the extent to which performance on the tests predicts performance on the job, both within and among the ethnic subgroups. Insight into the first issue can be gained only if individuals from all ethnic subgroups concerned are given the opportunity to perform the job. Insight into the second issue is readily gained by an examination of the performance on tests administered to members of the ethnic subgroups. Insight into the third issue can be gained only when test scores are correlated with job performance within ethnic subgroups—a condition predicated on the availability of job performance measures for reasonable numbers of ethnic subgroup members having held the job for an adequate length of time. Because the
ethnic group membership of job applicants ordinarily has not been identified by prospective employers, research on possible test discrimination has been unfeasible in most situations. Further, when selections have been based on test results, the data from such situations are unsuitable for sound research. Hence the limited research evidence related to the three issues in question.

The jcb classifications for the present project were selected because relatively high proportions of minority group personnel were holding the jobs, the jobs were considered as entry level positions with promotional possibilities, tests were already being used as screening devices in hiring, and because at a later point in time the employing organizations were willing to employ minority group members without using test results so as to permit a longitudinal study based on an uncontaminated sample.

GENERAL METHODOLOGY

The project reported herein was an outgrowth of a contract between the United States Equal Employment Opportunity Commission and the Colorado Civil Rights Commission. In the original application for financial assistance, the general objective of the project was stated as follows:

"To investigate the conformity of the testing and hiring procedures used by governmental agencies in Colorado with the EEOC Guidelines on Employment Testing.

The area in which intensive investigation is proposed is: Establishing whether or not the performance on the tests used in the selection of employees is related to on-the-job performance. If a relationship is found to exist, then to determine if the tests improperly discriminate against any ethnic group."

Upon approval of its application for financial assistance, the Colorado Civil Rights Commission subcontracted with Charles O. Neidt, Director, Human Factors Research Laboratory of Colorado State University, to collect and analyze appropriate data and report the results. In the subcontract with Colorado State University, the following steps were proposed:

"Step 1. Job analysis to determine job activities and characteristics considered essential for the successful performance of each job.

Step 2. Identification of criteria of successful performance for each job including supervisory ratings, supervisory rankings, training scores, personnel criteria, and turnover.

Step 3. Correlation of existing test data with performance criteria.

Step 4. Identification of additional testing devices for administration to present employees. These devices were to include both verbal and low verbal tests, as well as personal history inventory information. It was anticipated that at least 200 present employees would be administered the tests.

Step 5. Validation of additional devices through the computation of means and standard deviations for specific ethnic groups as well as the computations of coefficients of correlation with one or more criteria within each ethnic group.

Step 6. Longitudinal cross validation of selection techniques with an uncontaminated sample. On the basis of evidence from the two approaches, cross sectional and longitudinal, a specific set of recommendations for reducing possible discrimination through the use of tests in employment by governmental agencies in Colorado was to be developed."

Results of the project have been reported earlier in complete detail in two technical reports.\(^3\), \(^4\) The purpose of this report is to present a less technical and less detailed summary of the project.

**PROCEDURE**

To determine the specific activities involved in the jobs under consideration in the present project, interviews in which the job was discussed were held with at least two job incumbents, one immediate supervisor of the job and one department head. Available job descriptions were reviewed and incumbents were observed performing the job. From these sources a list of activities involved in performing the job was developed. These activities were then rated by incumbents and supervisors to identify the most important characteristics required for the successful performance of each job. It was noted that although the majority of the jobs being studied were at a relatively low level in the occupational hierarchy, they provided a variety of characteristics purportedly measured by employment test materials.

To obtain an indication of effectiveness on the job for the personnel involved in the project, five general classes of criteria were identified. These included (1) supervisory rankings, (2) supervisory merit ratings, (3) personnel criteria such as raises and promotions, (4) training scores, and (5) turnover. The supervisory ranking criterion was designated as the


measure of performance against which to compare test performance and the others were used to provide insight into how well members of the three ethnic groups performed on the job.

Review of the analyses of the jobs under consideration revealed considerable heterogeneity on both job activities and personal characteristics required for success. A battery of experimental tests was therefore assembled for administration in the cross sectional phase of the project which would tap a broad band of characteristics beyond those measured by tests currently in use. In addition, a biographical inventory was included in the experimental battery.

The total battery consisted of the following materials:

- **Oral Memory** - (ability to follow directions) - 10 minutes
- **Visual Memory** - (ability to recall material seen only briefly) - 6 minutes
- **Symbol Checking** - (ability to divide one's attention rapidly) - 12 minutes
- **Form Perception** - (ability to notice small differences among objects or symbols) - 14 minutes
- **General Mental Ability** - (ability to think abstractly) - 18 minutes
- **Verbal Ability** - (ability to define words) - 9 minutes
- **Personal History Information** - (educational, work, home and personal experiences) - 15 minutes

The total time for completing the experimental battery was one hour and twenty-five minutes.

All experimental tests were administered to present employees in classrooms, lunchrooms and conference rooms at their work sites. The number of employees tested at a given session ranged from three to sixty. Employees also indicated their ethnic group membership at this time. Job performance data were collected from employee files, from training reports,
or from the employee's supervisor. Preliminary analyses of the seven tests indicated that three of them were sufficiently indicative of job success to warrant further consideration. Arrangements were then made to use the Visual Memory Test, the Symbol Checking Test and the General Mental Ability Test in the longitudinal phase of the project.

In the longitudinal phase of the project, the three experimental tests were administered to applicants prior to their job interviews. These tests were not scored nor referred to in any way during the selection process. In fact, in no instance were the tests consulted until 90 days after the individual had applied for the job in question.

To assure that all applicants understood the purpose of the project, the following statement was presented to them prior to asking them to take the tests or indicate their ethnic group membership.

"The Colorado Civil Rights Commission and Colorado State University are making a study of how well applicants for civil service jobs perform on employment tests regardless of race, creed or color. We are therefore asking you to indicate your ethnic group in the space below. This information will be used for research purposes only. Although you are not required to give your ethnic background, you are urged to do so because in this way you can help eliminate possible discrimination in hiring."

JAMES F. REYNOLDS, DIRECTOR  CHARLES O. NEIDT, DIRECTOR
Colorado Civil Rights Commission  Human Factors Research Laboratory
Colorado State University

Throughout the two year project, only two individuals refused to participate.

From the foregoing description of procedures used in the project it can be noted that the project was conducted in two phases. Phase one involved administering an experimental test battery to present employees in selected jobs and relating their test score to their job performance. Phase two involved administering tests to job applicants and then following up those who were hired to assess their job performance at a later time.
Both phases contributed substantially to the results of the project.

RESULTS

Seven hundred fourteen employees participated in the cross sectional phase of the project. Of these employees, 120 were Negroes, 464 were Whites and 130 were Spanish-surname personnel. Seven hundred twenty-eight applicants, 187 of whom were later employed, participated in the longitudinal phase of the project. Of the longitudinal group tested, 96 were Negroes, 517 were White and 115 were Spanish-surname personnel. Of those employed, 28 were Negroes, 141 were Whites and 33 were Spanish-surname personnel.

Salient Findings

In general, although many specific findings were noted in the two technical reports based on the project, the three more pertinent findings as related to original questions posed in the project were as follows:

1. What differences in performance on selected jobs exist among three ethnic groups in Colorado?

In both the cross sectional phase of the project as well as the longitudinal phase, no evidence of differences in job performance among the three ethnic groups was found. This finding encompassed several jobs and several criteria of performance.

2. What differences in performance on selection devices exist among the three ethnic groups studied?

Although there were some exceptions and there was considerable overlapping of distributions, the average scores of Negroes and Spanish-surname personnel were lower than the average scores for Whites on most paper-and-pencil tests administered in the present project. Although the scores on low verbal materials tended to be more similar for the three
groups than the scores for highly verbal materials, differences still existed. Average oral interview ratings, however, were much more comparable for the three groups studied.

3. What differential relationships exist between performance on selection devices and performance on the job when employees in selected jobs are classified according to three ethnic groups?

The relationships between test scores and job performance tended to be highest for Negroes, next for Spanish-surname personnel, and lowest for Whites. Although the relationships between test score and job performance were not all high for all the tests investigated, more significant relationships were found for Negroes and Spanish-surname personnel than for Whites. This was especially true of the longitudinal phase of the project wherein the data were uncontaminated by cross sectional factors such as differences in length of employment, in differential change in the individual since time of application, etc.

In terms of the longitudinal prediction of job performance, prediction was most accurate within the Negro group, then within the Spanish-surname group, and least accurate within the White group. In fact, the relationship between test score and job performance for Whites was so low as to be useless for prediction purposes. In addition, the effectiveness of the oral interview ratings for predicting job performance was also low for all ethnic groups.

When the proportion of applicants completing each stage of the selection process was examined, it was found that the proportion of individuals from each ethnic group actually hired was remarkably close to the proportion of individuals from each ethnic group who had originally applied for the job. For example, among hospital attendant applicants, 24% were Negroes and among
the hospital attendants actually hired, 23% were Negroes. Among the applicants for intermediate clerk typist positions, 5% were Spanish-surname personnel and among those hired, 8% were Spanish-surname personnel. Among resident supervisor trainee applicants, 9% were Negro and 9% were ultimately employed.

It was noted, however, that time delays in the selection process (time lapse between application and interview, between interview and certification or between certification and final offer) tended to affect minority personnel adversely more than Whites.

When background characteristics of the three groups were studied, it was noted that the population of Negro applicants contained more individuals with high job performance potential than the other two groups in terms of percentage presently employed, increasingly higher salary on successive jobs and length of time on last job, even though this group tended to score lower on most tests than the other two groups.

It was further noted that the proportion of minority group personnel applying for higher level jobs was drastically lower than the proportion applying for lower level jobs. In fact, the absence of minority applicants in some of the higher level clerical classifications prevented the successful completion of the project for those classes (clerk stenographer, intermediate clerk stenographer, and senior clerk stenographer).

IMPLICATIONS

An axiom of the employment process is that, for selection to take place, there must be more applicants than position openings to be filled. Whenever this situation exists, every individual in the applicant population has a probability of being hired and a probability of not being hired. He also has a probability of succeeding if hired and a probability of being unsuccessful if hired. It is the task of the selection official
to so rank order the applicants that the probabilities of being hired parallel the probabilities of success for all applicants. In other words, if two individuals in the applicant population have equal probabilities of success on a job, they should also have equal probabilities of being hired. To the extent that individuals having equal probabilities of success on a job are assigned equal probabilities of being hired, then unfair discrimination has been avoided. To the extent that individuals having equal probabilities of success on a job are assigned unequal probabilities of being hired, then unfair discrimination exists.

To illustrate this concept, let it be supposed that there are two applicants for a position--a White man and a Black man. Let it also be supposed that both could perform the job equally well, if hired. For unfair discrimination to be avoided, both applicants must have an equal chance of being selected for the position; otherwise, unfair discrimination is introduced.

The project reported herein was designed to examine the selection processes used by two public employers in Colorado. There were two facets to the examination as shown by the following questions: (1) is there evidence of unfair discrimination in the selection processes used by the agencies, and (2) how accurate is the probability of success on the job as assigned to individuals in the applicant population at the time of application. In response to the first question, no evidence of unfair discrimination in terms of ethnic group membership was found. In response to the second question, it was found that it should be possible to improve selection accuracy by refining the process by which the "probability-of-success-on-the-job-if-hired" is determined. This refinement is based on the use of test results within ethnic groups and will require that the ethnic group membership of each applicant be known at the time of application. The
suggested refinement is described in the following paragraphs.

It will be recalled that there were no differences among the three ethnic groups in performance on the jobs studied. It will also be recalled that the relationship between test scores and on-the-job performance for Negroes was relatively high, for Spanish-surname personnel was intermediate, and for Whites was low, even though Negroes scored lowest on the tests as a group and Whites scored highest. There were also no differences among the interview ratings given during the "oral" examination, but the interview ratings had a low relationship to job performance for all three groups.

The interview ratings are ordinarily used to rank the applicants according to their probability of success on the job. This suggests that the probability of success on the job for Negroes and Spanish-surname personnel is not assessed as accurately with interviews as with tests, despite the fact that scores on the tests for the two minority groups were not as high as they were for Whites. Procedures are therefore needed which will more accurately identify the probability of job success for applicants than are currently available from the interview ratings alone. When applicants are listed on the basis of interview ratings, not all available information is being used to refine the listing. It is therefore suggested that more accurate selection could be obtained by modifying the position listing of Negroes and Spanish-surname personnel according to their test scores--since these scores relate well to job performance--and by leaving the positions of the Whites alone, since their scores do not relate well to job performance. In other words, the position on the list should reflect the relative probability of success on the job of the various applicants. By using the additional information obtained from the testing procedures, increased precision in the assignment of probabilities for success should be possible. This would mean rearranging the list so as to place high scoring Negroes
and Spanish-surname personnel (high scoring when compared with their own group) above some White applicants. The object of listing is to rank applicants in order of their predicted success on the job. If high scoring Negroes and Spanish-surname personnel have higher probabilities of job success than some Whites, then these individuals should be selected first.

To obtain an indication as to how this procedure might have affected the 66 persons employed by Denver General Hospital in the longitudinal phase of the present project, each minority group applicant's position in his ethnic group's distribution of one of the test scores was noted and a modified distribution for Negroes and for Spanish-surname personnel was constructed by adjusting weights to the interview ratings as follows:

<table>
<thead>
<tr>
<th>Test Score</th>
<th>Negro Weight</th>
<th>Spanish-Surname Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>80+</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>70-79</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>60-69</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

The smaller weights were assigned to the Spanish-surname group because of the lower relationship between their test score and their job performance. Whereas the relationship between the interview ratings and job performance for the total group without the adjustment was relatively low before the adjustment, this relationship increased substantially after the adjustment.

The change in position rankings of minority group members after the weighting was as follows: 11 of 16 Negroes employed were assigned weights; 1 was assigned 9 points and his performance category was 5 (highest possible on-the-job rating); 3 were assigned 6 points and their performance categories were 4, 4, and 3; 6 were assigned 3 points and the performance categories were 5, 4, 4, 3, 2 and 1. Thus it can be seen that 9 of the 11 individuals
assigned weights were average or above average performers after employment. Of these, 1 received 6 points (rating 3), 4 received 4 points (ratings of 5, 4, 3 and 3), and 5 received 2 points (ratings of 4, 3, 3, 3, and 1). In this instance only 1 individual was a below average performer.

A demonstration such as the foregoing tends to yield spuriously favorable results since the original relationship of test scores to job performance was determined on the same group as that to which the weights were applied. The procedure does have very meaningful possibilities, however, and will be discussed later as it might be applied in the practical setting of the agencies involved.

Whereas the foregoing procedure describes the interrelationships among job performance, selection test scores, and probability of success on the job, it does not provide an explanation for the reported conditions. Such an explanation can only be hypothesized from a logical analysis of the labor market and employment possibilities existing in Denver at the time the present project was conducted.

Most of the jobs studied in the present project (exclusive of clerical jobs) can be considered, at best, entry positions so far as the total job market is concerned. Both in terms of salary and job duties, they must be placed relatively low in the overall job hierarchy. This means that applicants would ordinarily apply for higher level jobs if such jobs were available to them, rather than lower-level jobs such as that of hospital attendant. It is here postulated that higher level jobs may have been available to the general White population at the time this project was conducted, but that higher level jobs were less available to minority group personnel. The minority group applicant population therefore contained many individuals who might have applied for higher level jobs.
had they had the opportunity to do so. On the other hand, it is postulated that the Whites did have such opportunities, thus leaving only less well qualified (even though higher scoring) White individuals in the applicant group. This, in turn, meant that it was possible to identify many "better qualified" Negroes and Spanish-surname personnel than Whites, and the presence of these individuals in the minority work group accounted for the higher relationships between test score and job performance found in the minority subgroups.

To obtain some indication of the types of job applicants in the minority group applicant pool, a comparison was made between the average score of the Denver Negro group and the average scores of some other Negro work groups who earlier had taken one of the tests investigated in the project (Matrices). These data were available from administrations of the test in other situations involving minority group personnel (one JOBS program and one training program). The score available for the other Negro groups and for the Denver Negro applicants for the job of hospital attendant were as follows:

<table>
<thead>
<tr>
<th>Origin of Data</th>
<th>Number of Cases</th>
<th>Job Classification</th>
<th>Arithmetic Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Francisco</td>
<td>10</td>
<td>Aircraft Cleaner</td>
<td>19.6</td>
</tr>
<tr>
<td>New York</td>
<td>10</td>
<td>Aircraft Cleaner</td>
<td>23.9</td>
</tr>
<tr>
<td>Newark</td>
<td>10</td>
<td>Aircraft Cleaner</td>
<td>21.6</td>
</tr>
<tr>
<td>Chicago</td>
<td>14</td>
<td>Aircraft Cleaner</td>
<td>19.6</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>7</td>
<td>Kitchen Helper</td>
<td>21.3</td>
</tr>
<tr>
<td>New York</td>
<td>10</td>
<td>Kitchen Helper</td>
<td>22.2</td>
</tr>
<tr>
<td>Denver</td>
<td>3</td>
<td>Kitchen Helper</td>
<td>29.3</td>
</tr>
<tr>
<td>San Francisco</td>
<td>2</td>
<td>Dining Service Helper</td>
<td>24.5</td>
</tr>
<tr>
<td>San Francisco</td>
<td>12</td>
<td>Kitchen Helper</td>
<td>18.8</td>
</tr>
<tr>
<td>Chicago</td>
<td>6</td>
<td>Janitor</td>
<td>15.0</td>
</tr>
<tr>
<td>San Francisco</td>
<td>5</td>
<td>Janitor</td>
<td>23.2</td>
</tr>
<tr>
<td>Jamaica</td>
<td>128</td>
<td>Mechanic</td>
<td>25.3</td>
</tr>
<tr>
<td>Total Denver Negro</td>
<td>74</td>
<td>Hospital Attendant Applicants</td>
<td>22.7</td>
</tr>
<tr>
<td>Employed Denver</td>
<td>17</td>
<td>Hospital Attendants</td>
<td>24.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Applicant</td>
<td></td>
</tr>
</tbody>
</table>
The Jamaicans were enrolled in a technical training program leading to positions as engine mechanics, were relatively young (early twenties) and had relatively good educational backgrounds (some were college graduates). The JOBS trainees were enrolled in training programs leading to entry-level positions in aircraft maintenance, janitorial and food service work. Most had not completed high school.

From the foregoing it is apparent that the Negro hospital attendant applicants were a relatively high scoring group in relation to other Negroes who had taken the test.

Additional evidence about the Denver applicant group is available from the background characteristics presented earlier. Here it will be recalled that the Negro group was characterized by upward mobility, job-wise, and by considerable stability in their former positions.

As a practical selection procedure, it is suggested that all applicants be interviewed first, rated on the basis of the interview, and then tested. Ethnic group membership could either be obtained voluntarily at the time of application or as the result of the interview (just as such membership in a minority group is now obtained by asking supervisors to report the number of minority group personnel under their supervision when surveys are made). On the basis of test scores interpreted according to minority group norms, the interview ratings could then be adjusted as they were in the foregoing demonstration. This procedure therefore becomes a concerted effort on the part of the prospective employer to assign the applicant with high probability of success to a realistic rank among the total applicant group.

The suggested procedure can be diagrammed schematically as follows:
Those minority individuals scoring in the lower portion of the distribution and passing the interview would have their ranking lowered accordingly. Scores of both the Whites and minority personnel would be tabulated on a bi-monthly basis so as to maintain a constant surveillance of the composition and nature of the applicant population. This suggested procedure will result in some minority group applicants being hired "ahead of" White applicants. Are these White individuals being unfairly discriminated against? The answer rests with the definition adopted for the term "Unfair discrimination." To the extent that a selection procedure increases the accuracy with which the probability of success on a future job is determined, then unfair discrimination is being avoided. The procedure here suggested simply modifies the assigned probabilities to be more nearly in line with the research evidence. As different research evidence is obtained, then the procedure must be changed.

The explanation of the findings of the project ties the results closely to the labor market within ethnic groups in the Denver area. It was postulated that because of some circumstance or other the Denver labor supply is more favorable within minority groups of applicants than within
the White group of applicants. As more higher level jobs become available to the Negroes and Spanish-surname personnel of the area, the composition of the entry-level job applicant groups will undoubtedly change. When this occurs, the selection procedure used to assign probabilities of success to each applicant in the total group must be changed. It is entirely conceivable that the relationship between test scores and job performance will become lower as this occurs. If it does, then the testing could be dropped altogether. On the other hand, if these entry level jobs are upgraded, or if some job crisis should occur (such as a major depression), and other jobs are disproportionately downgraded, then the White applicant population may change and this will require a change of selection procedure also. These possibilities simply underscore the critical need for additional research. As greater comprehension is gained into all facets of employee selection and placement, satisfaction of the employee and satisfaction of the employer are more nearly assured. Without the systematic collection of evidence, however, such comprehension is unlikely to result.