CONCURRENT VALIDATION OF THE COMPUTER PROGRAMMER APTITUDE BATTERY

THESIS

Presented to the Graduate Council of the North Texas State University in Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Dorsey W. Edwards, Jr. B.S.
Denton, Texas
August, 1981

Subjects were 34 computer programmers employed in a major computerized tax processing company. Scores in the Computer Programmer Aptitude Battery (CPAB) and ratings of each programmer's job performance by his immediate supervisor were obtained.

The purpose of the study was to validate a selection test. The relationship between the aptitude battery and performance evaluations was examined to evaluate the test's ability in predicting programming performance. Statistical treatment of data included Pearson product-moment correlations and a multiple linear regression analysis. The total test scores and several of the subtests were found to be significantly correlated with performance.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>iv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thesis</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Method</td>
<td>10</td>
</tr>
<tr>
<td>Subjects</td>
<td></td>
</tr>
<tr>
<td>Apparatus and Procedure</td>
<td></td>
</tr>
<tr>
<td>Predictors</td>
<td></td>
</tr>
<tr>
<td>Criteria</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td>14</td>
</tr>
<tr>
<td>Discussion</td>
<td>23</td>
</tr>
<tr>
<td>Appendix</td>
<td>25</td>
</tr>
<tr>
<td>References</td>
<td>30</td>
</tr>
</tbody>
</table>

iii
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Predictor Means, Standard Deviations, and $t$ Values for Computer Programmer Subgroups</td>
<td>16</td>
</tr>
<tr>
<td>2.</td>
<td>Intercorrelations among Predictors</td>
<td>17</td>
</tr>
<tr>
<td>3.</td>
<td>Criterion Means, Standard Deviations, and $t$ Values for Male/Females</td>
<td>18</td>
</tr>
<tr>
<td>4.</td>
<td>Criterion Means, Standard Deviations, and $t$ Values for Length of Employment</td>
<td>18</td>
</tr>
<tr>
<td>5.</td>
<td>Intercorrelation among Criterion Measures</td>
<td>20</td>
</tr>
<tr>
<td>6.</td>
<td>Predictor-Criterion Correlations</td>
<td>22</td>
</tr>
<tr>
<td>7.</td>
<td>Multiple Correlations of Summary Ratings with Grade-Point Average and Total Test Score</td>
<td>23</td>
</tr>
</tbody>
</table>
CONCURRENT VALIDATION OF THE COMPUTER PROGRAMMER APTITUDE BATTERY

Computer programming has become a major occupational field, and the need for proficient programmers has now far exceeded the supply of trained individuals. For this reason, the selection of individuals who show aptitude for programming is a most urgent problem.

The need for improved methods for the selection of computer programmers was recognized as early as 1952 (Rowan, 1952). Prior to this time, programmers were usually selected by means of personnel interviews and weighted application blanks. These methods combined with tests designed to measure programming aptitude are currently being used in selection procedures.

A major problem encountered with the use of tests was that many of them were "designed to yield measurements across the total population" (Palorno, 1974, p. 1). Unfortunately, only certain components have been effective for the specialized population of computer programmers. This fact established the need for tests of a more specific nature.

The validity of a test involves what the test measures and how well it does so. "Procedures for determining test validity are concerned with the relationships between performance on the test and other independently observable
facts about the behavior characteristics under considerations" (Anastasi, 1976). A strong relationship between the selection test and performance would serve to minimize employers' training expenses, reduce the failure rate of persons trained, defend the use of a selection device for purposes of EEO law, and improve productivity.

In a study of the effect of a valid test on the productivity of computer programmers, Schmidt and Hunter (1979), used decision theoretic equations, a newly developed technique to estimate the standard deviation of the dollar value of employee performance. The results supported the notion that large dollar amounts in increased productivity could be realized by increasing the validity of selection decisions in computer programmers. The study concluded that "the impact of valid selection procedures on work-force productivity is considerably greater than was originally believed" (p. 609).

The relatively new data processing field has not produced a shared description of a typical computer programmer. Educators and businessmen originally viewed programming as similar to problem solving. Programming then became closely associated with mathematics (Alspaugh, 1972). Mathematical knowledge is required to design computer programs, but the extent to which mathematical skill is necessary will depend on the type of program desired.
Alspaugh (1972) reported mathematical background had the greatest association between contributing factors and success in college computer programming courses. However, much of the total variance in classroom success was not accounted for, and there are "obviously other components of aptitude for computer programming yet to be identified" (p. 98).

Research has shown reasoning ability to be a factor in the ability to successfully master programming skills (McNamara & Hughes, 1961). The reasoning subtests of IBM's test of programming aptitude have been successful predictors of programming performance (IBM Manual, 1964). Additional evidence that reasoning ability is an important factor in programming ability has been provided by other researchers (Hollenbeck & McNamara, 1965; Howell, Vincent, & Gay, 1967; Palormo, 1980). The largest contributing factor of the Primary Mental Abilities Test (PMAT) for predicting job and training criteria was the reasoning section (Perry, 1964). Howell, Vincent, and Gay (1967) further reported that test of reasoning ability yielded the highest correlations when a simulated work sample was used as criterion.

Many other factors have been identified as influences on computer programming ability. Hollenbeck and McNamara (1965) reported several job factors which were repeatedly mentioned as being important: (a) ability to think logically, (b) mathematical ability, (c) ability to see a problem
through, (d) ability to work under pressure, (e) accuracy, (f) liking for detail, (g) a retentive memory, (h) ability to define the problem, and (i) the ability to work with people.

The validity research conducted prior to 1981 was mainly available only in the form of unpublished reports (McNamara & Hughes, 1961). Their review of the literature also revealed that several studies lacked important information such as sample size, correlation coefficient, and criterion used. Denelsky and McKee (1964) reported:

The number of published validation studies which have resulted from the use of selection tests seems disproportionately small in relation to the massive use of these tests as programmer selection devices. (p. 129)

Tests of computer programming aptitude have been divided into two main types: (a) paper-and-pencil tests, and (b) apparatus tests. The first apparatus test, the Problem-Solving Using Information (PSI) apparatus, was developed by John and Miller in 1957. The apparatus was designed specifically to measure problem-solving ability. Another apparatus developed by Bennett and Langmuir which followed the concept of the PSI apparatus was the Logical Analysis Device (LAD). McNamara and Hughes (1961) stated that

The LAD is used to test problem-solving ability by requiring the subject to light a center indicator light by the manipulation of nine lights in a circular display around the central light. (p. 46)
The Computer Usage Company Programmer Aptitude Test (CUCPAT) required the examinee to solve logical problems utilizing the lights of a 1401 IBM computer console (Hollenbeck & McNamara, 1965). The apparatus tests, although appearing to be job related, shared a number of disadvantages.

1. They required a trained individual who could administer and interpret the test.
2. Only one person could be tested at a time.
3. They were expensive to use.

Guion (1965) has questioned the practical value of the LAD, because it correlated highly with more economical paper-and-pencil tests.

Paper-and-pencil tests have had economic and administrative advantage when compared to the apparatus tests. However, they have demonstrated certain weaknesses also. Wolfe (1961) has pointed out limitations of aptitude tests intended for use with programmers. Factors such as "test-wiseness" and reading ability, although relating to test performance, had little or no relationship to job performance. Wolfe pointed out that certain tests such as the Programmer Aptitude Test are "more heavily loaded with mathematical applications than is necessary for business programming" (p. 270). The nature of speeded tests may have eliminated individuals whose slow, careful work habits were an asset to job effectiveness, according to Wolfe.

Predictors intended for use with the general population also have been widely used in an effort to predict programming
performance. Research has been conducted utilizing tests of interests (Bauer, Mehrens, & Vinsonhaler, 1968), personality (Alspaugh, 1972; IBM, 1964), general aptitude and ability (Bauer et al., 1968; Howell, et al., 1967; Katz, 1962) and biographical data (Alspaugh, 1972; Bauer et al., 1968; Howell et al., 1967). Tests such as the Color Naming Test (Howell et al., 1967) and the Brown-Carlsen Listening Comprehension (Upshall & Riland, 1958) have been used with varying degrees of success in predicting programmer performance.

The most widely used and researched test of programming aptitude has been IBM's Aptitude Test for Programmer Personnel (Howell et al., 1967). Originally developed in 1955, it was called the Programmer Aptitude Test (PAT) and Revised Programmer Aptitude Test (RPAT) in 1964. It became the Aptitude Test for Programming Personnel (ATPP) in 1965.

Data on the RPAT and ATPP have shown them to be "substantially related and that ATPP is as good or better than the RPAT in predicting class grades" (IBM, 1964, p. 7). Because the RPAT has been shown to be a valid predictor of job performance, the test authors suggested that the ATPP would also successfully predict job performance. The 1964 test manual reports a test score/ability to program rank order correlation of .49 (p. 11).

The author of the Aptitude Assessment Battery-Programmers (AABP) has claimed widespread administration of the test to thousands of individuals in companies, institutions, and
government agencies (Wolfe, 1970). Little evidence for its validity was available, however. A study by Denelsky and McKee (1974) reported the AABP to significantly predict both training and programming performance.

In the Computer Programmer Aptitude Battery (CPAB) manual, Palormo reported several validation studies of the CPAB. In 1972, a study of the CPAB's relationship to programming performance was conducted in a diversified paper products company. The CPAB total test and subtest scores correlated significantly with supervisory ratings of programmer trainee performance. All subtests were significant at the .01 level, except for letter series which was significant at the .05 level.

The CPAB was administered in 1978 (Palormo) to current computer programmers in a large computer service company. Each employee was rated by his/her immediate supervisor on a nine-job-factors performance rating. The performance rating was used as a criterion measure. All but one of the CPAB subtests, Verbal Meaning, had significant validities for predicting the performance criterion.

Palormo (1980) reported a recent concurrent validation study which was conducted in a large data processing company. The CPAB was given to 43 currently employed computer programmers and project leaders. The employees were rated by their immediate supervisors for their technical job performance. The results suggested that the number ability
subtest was the subtest most strongly associated with employees' technical job performance ratings. It was also noted that no significant difference was found with regard to sex or race.

Another study was carried out in 1978 (Palormo) which used 45 computer programmers who were currently employed in a Boston bank. The CPAB total and subtest scores were used to predict job performance ratings. Each employee was rated by his/her immediate supervisor on various job factors. The results showed the CPAB total score, as well as three of its subtests, was found to satisfactorily relate to job performance.

Standards for Education and Psychological Tests (1974) has specified methods employed for investigating the effectiveness of a test in predicting an individual's behavior in specified situations. For this purpose, performance on the test is checked against a criterion.

For many purposes, the most satisfactory type of criterion measure has been that based on follow-up records of actual job performance. Anastasi reported, "Most measures of job performance, although probably not representing ultimate criteria, at least provide good intermediate criteria for many testing purposes" (p. 144).

Negotiation and research by federal government enforcers, private and public employers, and public interest groups have produced the **Uniform Guidelines on Employee Selection**
Procedures (1978). The guidelines emphasize that criterion measures must represent unbiased measures of work behavior or performance. The importance of rating employees on objective job-oriented traits has been expressed by several researchers (Buel, 1970; Heir, 1970; Miner, 1968).

The possibility of bias or other contamination should be considered. A validation study of the CPAB in an eastern utility company reported the criterion measure (job-performance rating) was so contaminated by tenure that no inferences could be drawn from the study (Palormo, 1980). Other criterion problems have been rater reliability and restriction of range due to preselection of employees. "The reliability of a criterion measure should be estimated where feasible; estimates should be based on appropriate method" (Stanley, 1971, p. 63). "Reliability of a criterion measure need not be high, but there must be some reliability" (Thorndike, 1949, p. 65). Sanders and Peay (1974) warned that "rating forms that use ambiguous trait names on descriptions or require ratings on traits which are not observable lead to unreliable results" (p. 33). The reliability of the criterion in this study was estimated through test/retest reliability.

In view of previous research, a study to identify the relationships between predictors and performance seemed appropriate. The importance of generating validity coefficients has been pointed out by Schmidt, Rosenberg, and Hunter (1980). Their study noted "that small-sample validity studies
can contribute to cumulative knowledge and can make a useful contribution to the power of validity generalization models. Validity generalization models increase validity generalizability by allowing for the sources of error variance" (p. 653).

The test chosen for this study was the Computer Programmer Aptitude Battery. Veldman (Buros, 1972) reviewed the test and asserted that "the CPAB is probably the best device presently available for selection of computer programmers" (p. 1503). Veldman also remarked on the "obvious similarity between the various tests and on-the-job behavior" (p. 1503).

The purpose of this study was to investigate the relationship between the Computer Programmer Aptitude Battery and job performance in an organization which employs at least 20 to 30 new programmers a year.

Method

Subjects

Predictor and criterion data were collected on 25 female and 9 male programmers employed by a major computerized tax processing corporation. Individuals held various positions in the company's Systems Engineer career path.

Apparatus and Procedure

Predictors. The test selected was the Computer Programmer Aptitude Battery (CPAB), published by Science Research Associates. The test was selected on the basis of a review of the professional literature concerning the validity in
similar studies and an analysis of the job of systems engineers in this company.

The job analysis was conducted in order to gain an understanding of the requirements necessary for success as a computer programmer. The results of the job analysis were used to guide the selection of the aptitude battery which served as a predictor.

The method of job analysis employed in the study consisted of interviews and a task analysis questionnaire. The questionnaire focused on the time spent, importance, and difficulty level of job tasks and job activities. The questionnaire also covered person-related activities. Additional information about the job was provided through interviews which clarified questionnaire data and established task and activity cycles.

The job analysis described a list of ranked individual qualities and three categories of job-related activities: technical, cognitive, and interpersonal/communicative (see Appendix A). Through a review of the literature, the CAPB appeared to diagnose aptitude for a number of the technical and cognitive activities, as well as many of the individual qualities, identified in the job analysis.

The CPAB consists of five subtests designed to measure the following factors:
Subtest | Factor
---|---
Verbal Meaning | Communication skill
Reasoning | Ability to translate ideas and operations into mathematical notations
Letter Series | Abstract reasoning
Number Ability | Ability to estimate numerical computations
Diagramming | Ability to analyze problems and order the steps for solution in a logical sense

A score was obtained for each of the five subtests and for the test as a whole. College grade-point average was gathered from the incumbent's personnel file.

All test data were collected on self-scoring answer sheets by the same test administrator. Standardized conditions were maintained across subjects. All data were keypunched on separate cards, and all statistical analyses were carried out using standard computer programs at North Texas State University.

Criteria. The criteria against which the test battery was validated were performance ratings by the programmers' immediate supervisors on the job tasks and duties. The criteria used in the ratings were produced from a job analysis conducted by the personnel manager. Sanders and Peay recommend, "the list of traits to be evaluated, whether person-oriented, job-oriented, or both, should be determined from a thorough analysis of the jobs to be covered by the
evaluation" (p. 29). A list of standards of performance was submitted to four supervisors. These standards of performance were categorized into five major groups: (a) timeliness and accuracy; (b) technical ability; (c) self-improvement; (d) initiative, self-reliance, and responsibility; and (e) personal/professional skills. These five categories were each comprised of five to eight standards for evaluating programming performance (see Appendix B). The performance standards that were not considered relative to successful programming by at least three of the four supervisors were eliminated. The programmer supervisors, through consensus agreement, assigned relative weights to each category. The weight was based on each category's importance in a composite performance evaluation rating. The assigned weights were as follows: (a) timeliness and accuracy, (.30); (b) technical ability, (.20); (c) self-improvement, (.15); (d) initiative, self-reliance, and responsibility, (.15); and (e) personal/professional skills, (.20).

Ratings on the performance standards within each of the five categories were obtained by assigning the following values: (a) 1 = unsatisfactory; (b) 2 = does not consistently meet expectations; (c) 3 = consistently meets expectations; (d) 4 = consistently exceeds expectations; and (e) 5 = consistently outstanding. The category score was a subjective rating based on the performance standards. The rating for each category was multiplied by its assigned weight in order
to obtain the weighted rating for each category. The weighted ratings were added to obtain a single score having a value from 1 to 5 (this value corresponds to a scale of unsatisfactory to outstanding behavior).

Results

The five subtests' scores are coded by variable names: VM, (verbal meaning); R, (reasoning); LS, (letter series); NA, (number ability); and D, (diagramming). The total test score was coded as TS. The five categories (and summary rating) of performance criteria are computer coded as PE1, (timeliness and accuracy); PE2, (technical ability); PE3, (self-improvement); PE4, (initiative, self-reliance, and responsibility); PE5, (personal/professional skills); and PET, (total). Grade-point average is coded as GPA and represented by actual numerical GPA on a 4-point scale. In order to examine the effect of length of employment, tenure is coded as months of employment and is also collapsed into categories of less than 1 year and more than 1 year.

A Statistical Package for the Social Sciences program is run on the data. Means (M) and standard deviations (SD) for test scores and performance evaluation categories are computed for sex and length of employment. T tests of significance of difference between the means are performed. Intercorrelations among subtest scores are performed. Pearson product-moment intercorrelations of each performance rating (including totals) are performed. Pearson product-moment
correlations of each predictor score with each performance rating (including totals) are performed, and Pearson correlations of GPA as a predictor are performed with each performance rating (including totals). A multiple regression analysis is performed with the performance evaluation total against GPA and the total test score. Semipartial correlations of test scores with performance evaluation totals, with controls for tenure, are performed.

All data are analyzed for statistical significance. The aptitude battery is assessed for the degree to which it reflects success in computer programming.

Test performance is reviewed by total groups and by subgroups. In Table 1, predictor means, standard deviations, and t tests for each subgroup present in the sample are provided. In general, there appears to be no significant difference in scores for males and females or differences for length of employment.

In Table 2 are shown the intercorrelations among the predictor subtest scores. Correlations are, for the most part, moderate. This indicates that the subtests measure somewhat different traits. Tests showing the highest intercorrelations are tests which seem to measure a general reasoning factor (letter series, reasoning, and number ability). These results are, in general, consistent with reports in the CPAB tests manual (Palormo, 1980). Because CPAB total scores are a composite of all subscores, a fairly high
Table 1
Predictor Means, Standard Deviations, and $t$ Values for
Computer Programmer Subgroups

<table>
<thead>
<tr>
<th></th>
<th>Female N = 25</th>
<th>Male N = 9</th>
<th>Length of Employment</th>
<th>Total N = 34</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Verbal Meaning</td>
<td>21.1</td>
<td>4.5</td>
<td>20.1</td>
<td>6.8</td>
</tr>
<tr>
<td>Reasoning</td>
<td>14.1</td>
<td>4.9</td>
<td>11.1</td>
<td>4.5</td>
</tr>
<tr>
<td>Letter Series</td>
<td>15.6</td>
<td>3.9</td>
<td>14.4</td>
<td>5.4</td>
</tr>
<tr>
<td>Number Ability</td>
<td>13.5</td>
<td>4.9</td>
<td>15.2</td>
<td>4.2</td>
</tr>
<tr>
<td>Diagramming</td>
<td>29.9</td>
<td>4.5</td>
<td>30.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CPAB Total</td>
<td>94.4</td>
<td>14.5</td>
<td>90.8</td>
<td>18.1</td>
</tr>
</tbody>
</table>
test/subtest correlation is expected. All subtests are significantly correlated with the CPAB total at or greater than the .01 significance level which indicates good internal consistency.

Table 2
Intercorrelations among Predictors

<table>
<thead>
<tr>
<th>Test</th>
<th>Verbal Meaning</th>
<th>Reasoning</th>
<th>Letter Series</th>
<th>Number Ability</th>
<th>Diagramming</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Meaning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.2248</td>
</tr>
<tr>
<td>Reasoning</td>
<td>-.1192</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.1739</td>
</tr>
<tr>
<td>Letter Series</td>
<td>.1192</td>
<td>.3830**</td>
<td></td>
<td></td>
<td></td>
<td>-.0057</td>
</tr>
<tr>
<td>Number Ability</td>
<td>.0211</td>
<td>.6501**</td>
<td>.3813**</td>
<td></td>
<td></td>
<td>.0564</td>
</tr>
<tr>
<td>Diagramming</td>
<td>.2665</td>
<td>.4154**</td>
<td>.4033**</td>
<td>.3080*</td>
<td></td>
<td>-.2199</td>
</tr>
<tr>
<td>Total Score</td>
<td>.4342**</td>
<td>.7103**</td>
<td>.7053**</td>
<td>.7209**</td>
<td>.7154**</td>
<td>.0847</td>
</tr>
</tbody>
</table>

Note. *p .05       **p .01    N = 34

The stability of the criterion measure is estimated by test/retest reliability. The performance of the programmers was evaluated by all supervisors one month after initial review. The Pearson product-moment correlation between total test scores over 1 month is .90. The reliability of the rating is very high indicating similar responses over a period of time.

Criteria are, like the predictors, statistically analyzed to investigate any differences that could possibly be
attributed to either sex or tenure. The means, standard deviations, and \( t \) values for each criterion measure are shown in Tables 3 and 4. The mean summary rating and the mean timeliness and accuracy rating of programmers employed longer than 1 year is significantly higher than for programmers employed less than 1 year. Other differences are nonsignificant.

Table 3
Criterion Means, Standard Deviations, and \( t \) Values for Male/Female

<table>
<thead>
<tr>
<th>Variable</th>
<th>Female N = 25</th>
<th></th>
<th>Male N = 9</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Timeliness &amp; Accuracy</td>
<td>3.6</td>
<td>.43</td>
<td>3.4</td>
<td>.57</td>
<td>.83</td>
</tr>
<tr>
<td>Technical Ability</td>
<td>3.5</td>
<td>.73</td>
<td>3.4</td>
<td>.60</td>
<td>.34</td>
</tr>
<tr>
<td>Self-Improvement</td>
<td>3.2</td>
<td>.54</td>
<td>3.4</td>
<td>.43</td>
<td>-.96</td>
</tr>
<tr>
<td>Initiative/Self-Reliance</td>
<td>3.6</td>
<td>.68</td>
<td>3.5</td>
<td>.60</td>
<td>.48</td>
</tr>
<tr>
<td>Personal &amp; Professional Skills</td>
<td>3.6</td>
<td>.62</td>
<td>3.5</td>
<td>.40</td>
<td>.66</td>
</tr>
<tr>
<td>Performance Evaluation Total</td>
<td>3.5</td>
<td>.48</td>
<td>3.5</td>
<td>.46</td>
<td>.33</td>
</tr>
</tbody>
</table>

Table 4
Criterion Means, Standard Deviations, and \( t \) Values for Length of Employment

<table>
<thead>
<tr>
<th>Variable</th>
<th>&lt; 1 year N = 13</th>
<th></th>
<th>&gt; 1 year N = 21</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Timeliness &amp; Accuracy</td>
<td>3.3</td>
<td>.51</td>
<td>3.7</td>
<td>.39</td>
<td>-2.26*</td>
</tr>
<tr>
<td>Technical Ability</td>
<td>3.3</td>
<td>.86</td>
<td>3.6</td>
<td>.58</td>
<td>-1.16</td>
</tr>
<tr>
<td>Self-Improvement</td>
<td>3.1</td>
<td>.45</td>
<td>3.4</td>
<td>.52</td>
<td>-1.95</td>
</tr>
<tr>
<td>Initiative/Self-Reliance</td>
<td>3.3</td>
<td>.57</td>
<td>3.7</td>
<td>.66</td>
<td>-1.90</td>
</tr>
</tbody>
</table>
Quality of performance on specific duties appears to be closely associated with overall (mean) performance when ratings are subjected to a correlational analysis. In Table 5 the intercorrelations among all criterion measures are shown. Inspection of this table reveals that all criterion measures are substantially intercorrelated. Significant correlations indicate that if a programmer performs well in one category, he/she performs well in another. Another interpretation is that a certain amount of "halo" error might be evident in the ratings.

In order to discover relationships between test performance and performance on the job, Pearson product-moment correlations between all predictors and criteria were computed for the total sample. Results are entered in Table 6. The CPAB Reasoning, CPAB Number Ability, and CPAB Total show the best single correlations with the criteria. These scores correlate most highly with technical ability and evaluation total. This indicates the tests are good predictors of ability to understand programming problems and systems (technical
Table 5
Intercorrelation among Criterion Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Timeliness and Accuracy</th>
<th>Technical Ability</th>
<th>Self-Improvement</th>
<th>Initiative/ Self-Reliance</th>
<th>Personal and Performance Skills</th>
<th>Performance Evaluation Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeliness and Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical Ability</td>
<td>.6369*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Improvement</td>
<td>.5669*</td>
<td>.6660*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initiative/ Self-Reliance</td>
<td>.6810*</td>
<td>.6493*</td>
<td>.6291*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal and Professional Skills</td>
<td>.5705*</td>
<td>.4145*</td>
<td>.4790*</td>
<td>.5721*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance Evaluation Total</td>
<td>.8722*</td>
<td>.8311*</td>
<td>.7766*</td>
<td>.8488*</td>
<td>.7374*</td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .01  N = 34
ability) as well as overall programming performance. The Reasoning subtest appears to be the most accurate predictor of the performance categories. The Verbal Meaning, Letter Series, and Diagramming subtests of the CPAB did not demonstrate significant relationships with job performance. These results are, with the exception of the Diagramming Test, generally consistent with earlier findings relating CPAB scores to programming performance (Palormo, 1980). A possible explanation for the Diagramming result is that all programmers performed well on the Diagramming section, and the correlation dropped because of restriction of range (see Table 1).

The relationship of the performance evaluation total with the total test score and college GPA was analyzed using a multiple regression analysis. The subtests were not included in the analysis since the total score is a composite of all subtest scores. Also, due to the sample size of 34, it was decided to limit the predictors to a small number to avoid "shrinkage" of the multiple correlation coefficient.

In Table 7, multiple correlations between the predictors and criteria are shown. Although total test score and GPA are not highly correlated, the addition of GPA as a predictor does not significantly improve the correlation of predictors to criteria.
Table 6
Predictor-Criterion Correlations

<table>
<thead>
<tr>
<th>Criterion</th>
<th>GPA</th>
<th>Verbal Meaning</th>
<th>Reasoning</th>
<th>Letter Series</th>
<th>Number Ability</th>
<th>Diagramming</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeliness and Accuracy</td>
<td>.2253</td>
<td>-.0600</td>
<td>.3408*</td>
<td>.1544</td>
<td>.2899*</td>
<td>.0609</td>
<td>.2405</td>
</tr>
<tr>
<td>Technical Ability</td>
<td>-.0527</td>
<td>.1133</td>
<td>.4906**</td>
<td>.1263</td>
<td>.4196**</td>
<td>.3062*</td>
<td>.4486**</td>
</tr>
<tr>
<td>Self-Improvement</td>
<td>.1931</td>
<td>-.0234</td>
<td>.2879*</td>
<td>.0678</td>
<td>.3802**</td>
<td>.0275</td>
<td>.2295</td>
</tr>
<tr>
<td>Intiative/ Self-Reliance</td>
<td>.0838</td>
<td>-.0547</td>
<td>.2972*</td>
<td>.1327</td>
<td>.2274</td>
<td>-.2014</td>
<td>.0948</td>
</tr>
<tr>
<td>Personal and Professional Skills</td>
<td>.1674</td>
<td>-.0339</td>
<td>.2148</td>
<td>.2318</td>
<td>.1210</td>
<td>.0495</td>
<td>.1751</td>
</tr>
<tr>
<td>Performance Evaluation Total</td>
<td>.1598</td>
<td>-.0116</td>
<td>.4059**</td>
<td>.1844</td>
<td>.3583**</td>
<td>.0805</td>
<td>.3129*</td>
</tr>
</tbody>
</table>

Note.  *p < .05  **p < .01  N = 34
Table 7
Multiple Correlations of Summary Ratings with Grade-Point Average and Total Test Score

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Predictors</th>
<th>Multiple R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Evaluation Total</td>
<td>Total Score</td>
<td>.3129</td>
</tr>
<tr>
<td>Total</td>
<td>GPA</td>
<td>.3277</td>
</tr>
</tbody>
</table>

Because length of employment appears to be related to mean group performance on the criteria, a semipartial correlation is performed to remove the effects of tenure from the criterion. The zero-order correlation between total test score and performance evaluation total is .312. The semi-partial correlation controlling for tenure is .301. These results indicate no significant difference in correlations with the effect of tenure removed.

Discussion
Results of this study indicate that the job of computer programmer requires a number of varying skills and aptitudes. Results further indicate that commercially available tests can be helpful in predicting programmer performance. The Computer Programming Aptitude Battery demonstrates success in measuring the qualities necessary for successful performance as a programmer. It is suggested here that good reasoning ability and the ability to work with numbers are both necessary requirements when it comes to performing the job of computer programming.
As the computer industry evolves, the requirements necessary for success will also change. For this reason, it is recommended that further research be carried out using the CPAB and other tests shown to be successful in predicting performance.

Future research will be necessary for individuals and companies in the data processing industry to keep pace with the anticipated growth. Not only will the industry change on an overall basis, each company's needs and methods will change. Researchers in the selection process will be able to provide vital information for their company and the industry as a whole.
Appendix A

Systems Engineer Job Analysis Results

I. *Activities - ranked in descending magnitude (some overlap does occur)

A. Technical

1. Program and/or system maintenance/modification
2. Debugging
3. Coding
4. Testing
5. Program and/or system development
6. Documentation
7. Clerical

B. Interpersonal/Communicative

1. Maintaining customer relations
2. Working as a team member
3. Training others
4. Maintaining communications
5. Providing supervision and leadership
6. Conducting presentations

C. Cognitive

1. Satisfying decisions - making demands
2. Planning, organizing, scheduling
3. Problem definition
4. Assessing customers' needs - make recommendations
5. Research
II. *Individual Qualities* - ranked in descending magnitude
   A. Think and reason logically
   B. Work under pressure
   C. Depth-and-breadth perspective
   D. Self-confidence and enthusiasm
   E. Creative problem solving
   F. Adapt quickly to change (learning ability)
   G. Effectively understand/communicate technical and abstract concepts
   H. Respond well to criticism

III. **Time Cycles** - overview
   A. Client Technical Support - Basically bi-modal. Significant amplifications with fairly narrow variance in first year (during and after 1040 tax season). Second year amplification reduced about 2/3 with similar variance. First year amplification could be due in part to the fact that the contracts are not only new behaviors for SE but they are ground-laying as well. Note similarity to curve E - External Interpersonal Activity.

   B. Technical Performance - Cyclical curve that compliments A & E and varies with (and could be a result of) F. That is, F in large part could represent the project planning/designing stage while B represents the executions of these plans and designs. Logically these activities would vary inversely with A & E.

   C & D. Subordinate/Supervisor - Student/Teacher - Basically linear functions with similar slopes peaking at about 2.5 years. No real surprises here.

Note. *N = 9

**N = 6
Notes

Part I items, although job-related, run the range from objectifiable, technical items through predominately interpersonal/communication items to more cognitive items with a corresponding degree of difficulty in measurement.

Part II gets more into intrapersonal/cognitive variables that relate more to the job than classical personality traits and as a result may be more amenable to analysis under criterion-related rather than construct-related procedures.

Part III gives a visual picture of proportion of time spent in three major areas of activity. This should help us weigh the relative importance of the overall findings.
Appendix B
Performance Criteria System Engineer

The candidate's performance will be determined by his/her success in:

1. Ability to complete projects and provide technical assistance timely and accurately (30)
   - Able to prioritize and organize workload
   - Notifies supervisor when deadlines are slipping
   - Manages time wisely to work within external limitations
   - Willing to work hours necessary to complete projects on time
   - Controls and adapts to interruptions, changes and disorders tactfully; does not let others waste their time
   - Responds to client and Fast-Tax personnel's technical inquiries timely and accurately

2. Technical Ability (20)
   - Understands tax law in order to analyze and understand accounting definitions and requirements of finished product
   - Demonstrates grasp of PL1 and Fast-Tax system
   - Pays attention to detail; attempts to search out answers to program problems but knows when to ask questions
   - Demonstrates ability to understand and grasp complex programming problems

3. Self-Improvement (15)
   - Keeps informed of changes in area of expertise
   - Expands knowledge of data processing
   - Develops and presents to others new ideas and solutions
   - Grasps new concepts, approaches, or systems

4. Initiative, Self-Reliance and Responsibility (15)
   - Takes action when appropriate on own initiative
   - Asks for additional work when workload is slack
   - Knows when authority is being overstepped, obtains proper authorization
   - Takes responsibility for projects from analysis, through testing to documentation
5. Personal and Professional Skills (20)

- Displays a positive attitude toward Fast-Tax and takes pride in profession and work
- Is courteous to clients and others contacted
- Tactfully resolves conflict and/or problem situations with others
- Cooperative and courteous
- Considerate of other people's time
- Resolves competing priorities and still maintains a good working relationship with those involved
- Strives to improve from constructive criticism
- Consistently adheres to reasonable hours of work as determined by your manager
- Observes approved policies and procedures
- Attempts to resolve own personal conflicts with immediate supervisor; if still dissatisfied, will discuss problems with manager
References


Perry, D. K. Training and job performance: Validities of programmer trainee selection variables. Santa Monica,


