DOCUMENT RESUME

ED 068 546 TM 002 027

TITLE Cross-Validation of the General Aptitude Test Battery and Development of a Weighted Application Blank for Computer Technology Trainee 828.XX.
INSTITUTION Manpower Administration (DOL), Washington, D.C. U.S. Training and Employment Service.
REPORT NO S-415; USES-TRR-19a
PUB DATE Jan 69
NOTE 34p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Aptitude Tests; *Computer Science; *Cutting Scores; Evaluation Criteria; Job Applicants; *Job Skills; Norms; Occupational Guidance; *Personnel Evaluation; Test Reliability; Test Validity; Trainees
IDENTIFIERS GATB; *General Aptitude Test Battery

ABSTRACT

The United States Training and Employment Service General Aptitude Test Battery (GATB), first published in 1947, has been included in a continuing program of research to validate the tests against success in many different occupations. The GATB consists of 12 tests which measure nine aptitudes: General Learning Ability; Verbal Aptitude; Numerical Aptitude; Spatial Aptitude; Form Perception; Clerical Perception; Motor Coordination; Finger Dexterity; and Manual Dexterity. The aptitude scores are standard scores with 100 as the average for the general working population, and a standard deviation of 20. Occupational norms are established in terms of minimum qualifying scores for each of the significant aptitude measures which, when combined, predict job performance. Cutting scores are set only for those aptitudes which aid in predicting the performance of the job duties of the experimental sample. The GATB norms described are appropriate only for jobs with content similar to that shown in the job description presented in this report. A description of the validation sample is included. (AG)

January 1969 United States Employment Service Test Research Report No. 19a

Cross-Validation of the General Aptitude Test Battery and Development of a Weighted Application Blank for Computer Technology Trainee 828.XX

S-415

U.S. DEPARTMENT OF LABOR

MANPOWER ADMINISTRATION

ED 068546

USES Test Research Report No. 19a

Cross-Validation of the General Aptitude Test Battery and Development of a Weighted Application Blank

For

Computer Technology Trainee 828.XX

S-415

U.S. DEPARTMENT OF LABOR

MANPOWER ADMINISTRATION

BUREAU OF EMPLOYMENT SECURITY

U.S. EMPLOYMENT SERVICE

January 1969

FOREWORD

Extensive research conducted under the Federal-State cooperative test research program in the Employment Service has led to the development of many tools useful in vocational counseling and placement. These tools include aptitude tests, proficiency tests, and non-cognitive measures based on instruments such as interest inventories and biographical information blanks.

The purpose of this series of reports is to provide results of significant test research projects as they are completed. These reports will be of interest to users of the tests and to test research personnel in other organizations.

U.S. Employment Service

CROSS-VALIDATION OF USES APTITUDE TEST BATTERY

For

Computer Technology Trainee 828.XX[a]

S-415

This report describes research undertaken for the purpose of cross-validating General Aptitude Test Battery (GATB) norms for Computer Technology Trainee 828.XX.[b] Further analysis was undertaken to (1) cross-validate the previously developed multiple regression equation and compare this research technique with the multiple cut-off method, (2) continue the evaluation of the SRA Tiffin Adaptability Test and the California Achievement Test-Math-Part B as predictors of successful training completion, and (3) develop a weighted application blank using a portion of the original sample, validate it on the remaining portion, and cross-validate it on the entire second sample. The following norms were established in the original study as test battery S-415:

GATB Aptitude	Minimum Acceptable GATB Scores
G - General Learning Ability	110
N - Numerical Aptitude	95
S - Spatial Aptitude	100

RESEARCH SUMMARY

Sample:

173 male students enrolled in a Computer Technology course at Control Data Institute, Minneapolis, Minnesota.

Table 1 gives the quantitative analysis of the means, standard deviations, and correlations with the criterion for the GATB aptitudes, age, and education. For this analysis the entire usable sample, rather than the cross-validation sample, was considered in order to assure that the exclusion of the band group did not alter the regression line.

[a]Research conducted by Minnesota State Employment Service in cooperation with Control Data Institute.

[b]A description of the research included in the original study may be found in the United States Employment Service Test Research Report No. 19.

TABLE 1

Means, Standard Deviations (SD), Ranges, and Pearson Product-Moment Correlations with the Criterion (r) for Age, Education, and GATB Aptitudes for the Cross-Validation Study

N = 241

	Mean	SD	Range	r
Age (years)	23.9	5.4	17-47	.270**
Education (years)	12.5	.9	10-16	.114
G-General Learning Ability	118.3	12.6	83-161	.451**
V-Verbal Aptitude	108.2	11.5	68-141	.436**
N-Numerical Aptitude	115.0	13.3	82-156	.313**
S-Spatial Aptitude	121.9	14.9	71-163	.266**
P-Form Perception	118.8	17.0	72-174	.092
Q-Clerical Perception	116.9	13.4	87-174	.216**
K-Motor Coordination	102.2	16.3	41-159	.013
F-Finger Dexterity	103.8	17.9	55-156	.101
M-Manual Dexterity	110.7	19.5	47-163	.136

**Significant at the .01 level

Criterion:

Rounded average of all unit exams

Design:

Longitudinal (test data were collected on the second day of training and criterion data were collected at the end of the ten-month training program).

Principal Activities:

A summary of the course and the syllabus are presented in the Appendix to Report No. 19.

Predictive Validity:

Phi Coefficient using Yates' correction = .397 ($P/2 < .0005$)

Effectiveness of Norms:

Only 75% of the non-test-selected students used for this study were good students; if they had been test-selected with the S-415 norms, 85% would have been good students. 25% of the non-test-selected students used for this study were poor students; if they had been test-selected with the S-415 norms, only 15% would have been poor students. The effectiveness of the norms when applied to this independent sample is shown in Table 2.

TABLE 2

Effectiveness of S-415 Norms with the Cross-Validation Study

	Without Tests	With Tests
Good Students	75%	85%
Poor Students	25%	15%

TABLE 3

Predictive Validity of Test Norms
(G-110, N-95, S-100)
Cross-Validation Sample

	Nonqualifying Test Scores	Qualifying Test Scores	Total
Good Students	18	111	129
Poor Students	24	20	44
Total	42	131	173

Phi Coefficient* (ϕ) = .397
Significance Level = $P/2 < .0005$

Chi Square* (χ^2_y) = 27.2

SAMPLE DESCRIPTION

The final cross-validation sample consisted of 173 students enrolled in classes 9 through 13. Classes 2 through 8 had previously been used for the original study research, and class 1 was eliminated because testing had not been undertaken until midway through their training program. By this time a majority of the poor students had already terminated, and for the remaining students there was the possibility of elevated aptitude scores resulting from the intensified training. Table 4 shows the number of students from each class included in the final sample.

TABLE 4

Sample Breakdown Per Class

Class	9	10	11	12	13	Total
Number	40	34	36	36	27	173

* Using Yates' correction

The procedure developed in the original data analysis of excluding a "band group" of marginal status students was continued with this cross-validation sample. There were two possible methods of selecting the band group: (1) by maintaining the same two cutting scores on the criterion as used in the original study of above 75 for the good group and below 60 for the poor group and eliminating those students inbetween or (2) by employing the procedure developed with the original study of determining the percentage of poor students, doubling this percentage for the good group and excluding the middle percent. It was felt that the first method was preferable because it maintained the same achievement levels for the two samples. Therefore, the actual good to poor ratio for the cross-validation sample was 75% - 25% instead of 66% - 34% as in the original study.

As with the first study, there were a number of students who terminated the training for reasons other than scholastic failure and could not be included in the final sample. Table 5 summarizes the breakdown of the tested sample. The mean, standard deviation, and range on the criterion for each subsample is given in Table 6.

TABLE 5

Tested Sample Breakdown

Identification	N
Total Number Tested	336
Number not Usable	95
Total Usable Sample	241
Number Included in Cross-Validation Sample	173
Good Students - 129	
Poor Students - 44	
Number in Band	68

TABLE 6

Mean, Standard Deviation, and Range on the Criterion for Each Subsample

	N	M	SD	Range
Good Students	129	84.0	5.7	76-97
Poor Students	44	50.4	7.5	28-59
Band Group	68	71.1	3.3	63-75
All Graduates	197	79.6	7.9	63-97

A comparison of the difference in means between the SATB and cross-validation samples for each variable was made using the critical ratio test for independent samples.[1] Although there was a trend toward

[1] Garrett, H. E., Statistics in Psychology and Education, (New York: David McKay Co., 1964), pp. 214-215.

higher means in the cross- validation sample, these differences were significant only for the factors of age, N, F, unit exam average, and Tiffin. Thus, it would appear that the two samples are roughly comparable but that gradually the Institute is striving to select older, higher ability trainees.

EXPERIMENTAL TEST BATTERY

All twelve parts of the General Aptitude Test Battery, B-1002B were administered between November 11, 1966, and June 20, 1967.

CRITERION

The criterion remained identical to the one employed in the SATB development with the minor exception that beginning with Class 11, the unit exams constituted 100% of the final grade for Phase III. Previously, a rating scale evaluation for each student had been completed by the instructor at the time of the examination in Phase III and scored in such a manner that 25% of the exam grade reflected the rating scale score. This procedure has been discontinued by Control Data Institute because it was felt that it only tended to lower the grades of the top students and thereby lessened the differences between good and poor students.

Criterion Score Distribution for Cross-Validation Sample:

Possible Range	0-100
Actual Range	28-97
Mean	75.5
Standard Deviation	15.9

CROSS-VALIDATION OF THE NORMS

A comparison of the phi coefficient from the SATB development with the cross-validation phi, as shown in Table 7 indicates very minor loss in the predictive validity of the norms when applied to a new sample.

TABLE 7

Comparison of the Phi Coefficient of the SATB Sample with the Cross-Validation Sample

SATB corrected $\phi = .416$

Cross-Validation corrected $\phi = .397$

Due to these results, it no longer seemed necessary to consider other methods of deriving valid norms, such as combining both samples or drawing one line of demarcation between the high and low criterion groups.

MULTIPLE REGRESSION ANALYSIS

As part of the data analysis in the Phase I SATB development study, a comparison was made between the research techniques of multiple regression and multiple cut-off. Anastasi, discussing the comparative advantages and disadvantages of each procedure, suggests a combination of both methods such that the multiple cut-off is applied first and then the multiple regression equation is used to compute a predicted criterion score for each individual passing the cut-off norms.[2] Following this proposal, the Wherry Doolittle Test Selection method was chosen as the statistical technique for obtaining the regression equation. The resulting raw score prediction formula was:

$$Y' = .376X_G + .250X_N + .139X_S - 18.001$$

where

Y' = Predicted unit examination average

X_G = Score for aptitude G

X_N = Score for aptitude N

X_S = Score for aptitude S

When the predicted correlation between the regression equation and the criterion ($\overline{R} = .462$) was compared with the actual obtained correlation ($r = .470$), there was very close agreement. Next, the criterion scores predicted from the multiple regression equation were analyzed in a manner similar to standard SATB procedures in order to determine the best cutting score and phi coefficient. The computed corrected phi coefficient value was .393.

There were two methods available to determine the validity of the multiple regression aptitude beta weights when applied to the cross-validation sample. First, after computing the predicted criterion scores from the regression equation, a correlation was run between the predicted and actual criterion values on the cross-validation sample, resulting in an r of .451. Although this value was somewhat below the correlation of .470 obtained with the SATB sample, it does indicate a high degree of validity for the beta weights. Second, by using the cutting score of 67 (which was determined on the validation sample to have the best selective efficiency) and setting up a four-fold table it was possible to check the regression equation validity on the cross-validation sample another way. This information is located in Table 8.

---

[2] Anastasi, Anne, Psychological Testing, (New York: The Macmillan Co., 1961), pp. 172-177.

TABLE 8

Cross-Validation of the Multiple Regression Cutting Score of 67

N = 173

	Nonqualifying Test Scores	Qualifying Test Scores	Total
Good Students	15	114	129
Poor Students	21	23	44
Total	36	137	173

Corrected ϕ = .371 P/2 $<$ .0005
Corrected χ^2_y = 23.8

Again, as with the SATB norms, the phi coefficient has decreased for the cross-validation sample, but the magnitude of this decrease is such that there has been minimal loss in the predictive validity of the regression equation. A summary of the results of the two methods for determining the validity of the multiple regression equation is presented in Table 9.

TABLE 9

Summary of the Validity of the Multiple Regression Equation

Obtained correlation between equation and criterion for SATB sample	r = .470
Obtained correlation between equation and criterion for cross-validation sample	r = .451
Phi coefficient for SATB sample	ϕ = .393
Phi coefficient for cross-validation sample	ϕ = .371

One other evaluation made in the Phase I study was to add the multiple regression cutting score to the aptitude cutting scores in order to determine the selective efficiency of both techniques combined. This resulted in greater differentiation between successful and unsuccessful trainees than either method used alone. (Phi coefficient = .426). Table 10 shows the phi coefficient value when this same procedure was tried for the cross-validation sample.

TABLE 10

Selective Efficiency of the SATB and Multiple Regression Combined for Cross-Validation Sample

N = 173

	Nonqualifying Test Scores	Qualifying Test Scores	Total
Good Students	20	109	129
Poor Students	24	20	44
Total	44	129	173

Corrected $\phi = .375$ $\quad$ $P/2 < .0005$
Corrected $\chi^2_y = 24.4$

Therefore, for the cross-validation sample, a combination of both methods is superior to just the multiple regression equation, but the SATB alone provides even greater selective efficiency.

Having developed, validated, and cross-validated both research procedures on two samples of computer technology trainees, several tentative conclusions now seem justified: (1) initially, both methods provide good selective efficiency for the sample, with the multiple cut-off providing slightly greater differentiation, (2) for the cross-validation sample the loss in validity, using the phi coefficient, for both methods is slight and approximately equal, (3) the validity of the regression equation, as measured by the correlation between the equation and the criterion for the two samples is high, and (4) treating the multiple regression equation as a fourth cutting score and combining it with the SATB is of questionable value. The Minnesota agency is interested in pursuing the possibilities of the multiple regression technique further in other studies by considering differing levels of the regression score as predictors of occupational success. This would provide knowledge similar to that obtained for the multiple cut-off method from United States Employment Service Test Research Report No. 12. If the regression score could be used to predict level of success, it could then be applied as an additional measure in instances when it is necessary to select among those applicants passing the SATB norms.

ANALYSIS OF THE TIFFIN AND CALIFORNIA

Two commercial tests, the California Achievement Test - Math - Part B and the SRA Tiffin Adaptability Test, have been used by Control Data Institute as guidelines in the selection of trainees. The California Achievement Tests, advanced form, are designed as a scholastic battery for grades 9 thru 14 measuring reading, mathematics, and language. Part B of the mathematics section, administered by the Institute, covers the areas of symbols, rules, equations, and the use of negative numbers in equations. There are 25 items in Part B arranged in a five-response, multiple choice format with a ten minute time limit.

A variety of questions pertaining to number series, vocabulary, word problems, and general knowledge comprise the 35-item SRA Tiffin Adaptability Test. As described by the authors, the Tiffin is a fifteen minute power test designed as an employment aid to measure mental adaptability or mental alertness. In the manual, a suggested range of scores for those jobs at the level of a technician is 24 to 29 points; however the authors state the importance of deriving local norms for a given occupation.

Table 11 gives the correlations between (1) the Tiffin and the California and (2) each of these two commercial tests with the variables of age, education, unit exam scores, and GATB aptitudes for the cross-validation sample. These correlations were based on a sample size of 134 trainees, which is the total number of individuals of the 173 students in the cross-validation study who took both commercial tests. In general, these correlations are lower than those obtained in the original study.

TABLE 11

Correlations between the Tiffin and California and Between Each of the Tests with Age, Education, Unit Exam Scores, and GATB Aptitudes for the Cross-Validation Sample

N = 134

Factor	California	Tiffin
Tiffin	.415**	----
Age	-.238**	.102
Education	.237**	.152
Unit Exam Scores	.280**	.538**
G	.385**	.553**
V	.323**	.581**
N	.361**	.357**
S	.205*	.367**
P	.113	.102
Q	.113	.298**
K	.044	-.071
F	-.048	.081
M	-.027	.082

* Significant at the..05 level
** Significant at the .01 level

Applying the cut-off of 18 on the Tiffin and 19 on the California, as developed in the original study, the phi coefficient and chi square were computed for the cross-validation sample. This information is found in Table 12.

TABLE 12

Selective Efficiency of Tiffin-18 and California-19 on the Cross-Validation Sample

N = 134

	Nonqualifying Test Scores	Qualifying Test Scores	Total
Good Students	35	66	101
Poor Students	22	11	33
Total	57	77	134

Corrected $\phi = .262$ $P/2 < .005$
Corrected $\chi^2_y = 9.2$

By comparing this phi coefficient with the value of .407 obtained in the original study, it may be seen that the validity of these norms has been substantially reduced with a new sample. There is now a significant difference between the predictive validity of the SATB norms and the Tiffin and California as is shown in Table 13.

TABLE 13

Comparison of Phi Coefficients for the SATB Norms and the Tiffin and California

Validation Sample	Cross-Validation Sample
SATB norms $\phi = .416$	SATB norms $\phi = .397$
Tiffin and California $\phi = .407$	Tiffin and California $\phi = .262$

Therefore, it is recommended that, as a measure of training ability, the Specific Aptitude Test Battery norms of G-110, N-95, and S-100 be used in place of the Tiffin and California for predicting successful completion of computer technology training.

DEVELOPMENT OF A WEIGHTED APPLICATION BLANK

Recognizing the need to explore predictors of occupational success other than ability, an attempt was made in this study to develop a weighted application blank using the procedures suggested by England.[3] The following items were listed on the application blank at Control Data Institute and were therefore considered in the analysis of the weighted application blank (WAB):

1. Age, height, and weight
2. Marital status and number of children
3. Father's occupation
4. Year graduated from high school, class standing, and favorite subject
5. High school courses in algebra, trigonometry, geometry, physics, and electricity or radio
6. Number of years of college
7. College course work in mathematics, physics, chemistry, electrical engineering, mechanical engineering, and other engineering subjects
8. Practical experience in electronics
9. Presently employed
10. Desire part-time work with Control Data Corporation

Selection of the Sample Groups:

Since those students leaving for nonacademic reasons could not be considered, the potential sample for the development and validation of the WAB consisted of all students in classes 1 through 9 who either failed or completed training at the Institute. Because GATB test scores were not involved, class 1 could be included in this data analysis. Class 9 was also added to the original SATB sample in order to obtain a sufficiently large low criterion group. With these two extra classes, there was a total of 82 students who had a unit exam average below 60 points and had completed an application blank. These students formed the low criterion group.

In order to maximize the differentiating factors between successful and unsuccessful trainees, only those 90 students who earned an exam average above 82 points were included in the high criterion group. This resulted in an approximately equal number of students in the high and low criterion groups. Each criterion group was then dichotomized into a weighting group and a holdout group, using the suggested two to one ratio. The weighting group was used to identify and weight differentiating items, and the holdout group provided a means of evaluating the effectiveness of the WAB and determining a cutting score. A table of random numbers was used for the dichotomization.

---

[3] England, G. W., Development and Use of Weighted Application Blanks, (Dubuque, Iowa: Wm. C. Brown Co., 1961).

Selection and Weighting of Items:

Most of the items were such that the responses could easily be divided into two or three categories; however, other items were either continuous in nature or had numerous response categories. For the continuous variables of age, height, weight, and number of years since high school graduation a frequency distribution was first constructed for the combined weighting groups and then the distribution was divided into three or four equal frequency intervals. Father's occupation was split into the categories of professional, clerical, farming, skilled trades, and unskilled labor. Favorite subject was divided into math courses, science courses, other college related subjects, and business or shop courses. All other variables had natural breaking points.

In addition to the application items already listed, two more items were also considered - the number of related high school courses and the number of related college courses. In order to determine item weights, it was necessary to compute the percentage of high criterion students in each response category, the percentage of low criterion students in each response category, and the percent difference. By using appropriate tables,[4] the percent difference can be converted into a net weight and then to an assigned weight. Assigned weights are simply a scoring convenience since they eliminate negative weights and reduce the numerical value of the net weights, thus making it easier to add weights together. Possible values for assigned weights are zero, one, and two.

After determining the initial frequency distributions, percentages, and weights, certain adjustments were made in combining and eliminating items. The number of children and having taken electrical engineering, mechanical engineering, or other engineering subjects were all eliminated because there were too few responses in each category to make them valid items. It was felt that item weights which were developed on just a few responses were questionable and subject to considerable fluctuation and should therefore not be considered. The number of years since high school graduation was also eliminated because in reality it measured the same factor as age, thereby providing a double advantage or disadvantage for only one item.

Several factors influenced the decision to combine the three differentiating items - studied college math, studied college physics, and number of related college courses taken-into one item with the following response categories, (1) neither math nor physics, (2) math but no physics, and (3) physics. The primary reason for the combination was because there was a disproportionate number of questions pertaining to this area considering the small portion of students who had any college experience. Also, several of these items seemed to be measuring the same area. Generally, those students who had taken physics had also taken math and had been enrolled longer with more courses in related subjects. An analysis of the items suggested combining them into the format stated above.

---

[4]Ibid., pp. 23-25.

Preliminary research indicated that having taken a college physics course was really the differentiating factor between the high and low criterion groups since practically all of those students taking physics had also studied college math. When used as separate items, college math was given a spuriously high weight because in most cases it was more accurately measuring whether a student took physics.

According to England's procedure, the item of college math or physics is a "contingent" item because answering it depends on whether the student attended college. If these items were scored the usual way, a student would be penalized once for not attending college and a second time for not taking college math or physics. However, the recommended scoring procedure is to record an assigned weight of one to all persons unable to respond to such contingent items. Being midway in the assigned weight range, a score of one offers neither an advantage nor a disadvantage to the individual.

Presented in Table 14 is a list of all items which were included or eliminated from the final WAB as well as the possible response categories. (A copy of the final WAB is shown in the Appendix.)

TABLE 14

Weighted Application Blank Items

Items Selected for Final WAB:

1. High school class standing (upper third, middle third, lower third)
2. Favorite subject in high school (math, science, other college related course, business or shop course)
3. Had high school trigonometry (yes, no)
4. Had high school geometry (yes, no)
5. Had high school physics (yes, no)
6. Number of completed years of college (none, less than one, one, two or more)
7. Practical work experience in electronics (yes, no)
8. Number of related high school courses (one, two, three, four, five)
9. Father's occupation (professional, clerical, farming, skilled, unskilled)
10. Age (17-18, 19-20, 21-23, 24+)
11. College course work (neither math nor physics, math but no physics, physics, not applicable)

Items Which Did Not Differentiate between High and Low Criterion Groups:

12. Marital Status (single, married)
13. Had high school algebra (yes, no)
14. Had high school electricity or radio course (yes, no)
15. Had college chemistry (yes, no, not applicable)
16. Presently employed (yes, no)
17. Desire part-time work with Control Data Corporation (yes, no)
18. Height (0-69, 70-71, 72+ inches)
19. Weight (0-145, 150-165, 170+ pounds)

Items Which Had Too Few Responses to be Used:

20. Number of children (zero, one, two, three or more, not applicable)
21. Had college electrical engineering subjects (yes, no, not applicable)
22. Had college mechanical engineering subjects (yes, no, not applicable)
23. Had other college engineering subjects (yes, no, not applicable)

Items Which Were Combined:

24. Had college math (yes, no, not applicable). Combined to form item #11.
25. Had college physics (yes, no, not applicable). Combined to form item #11.

Items Deleted:

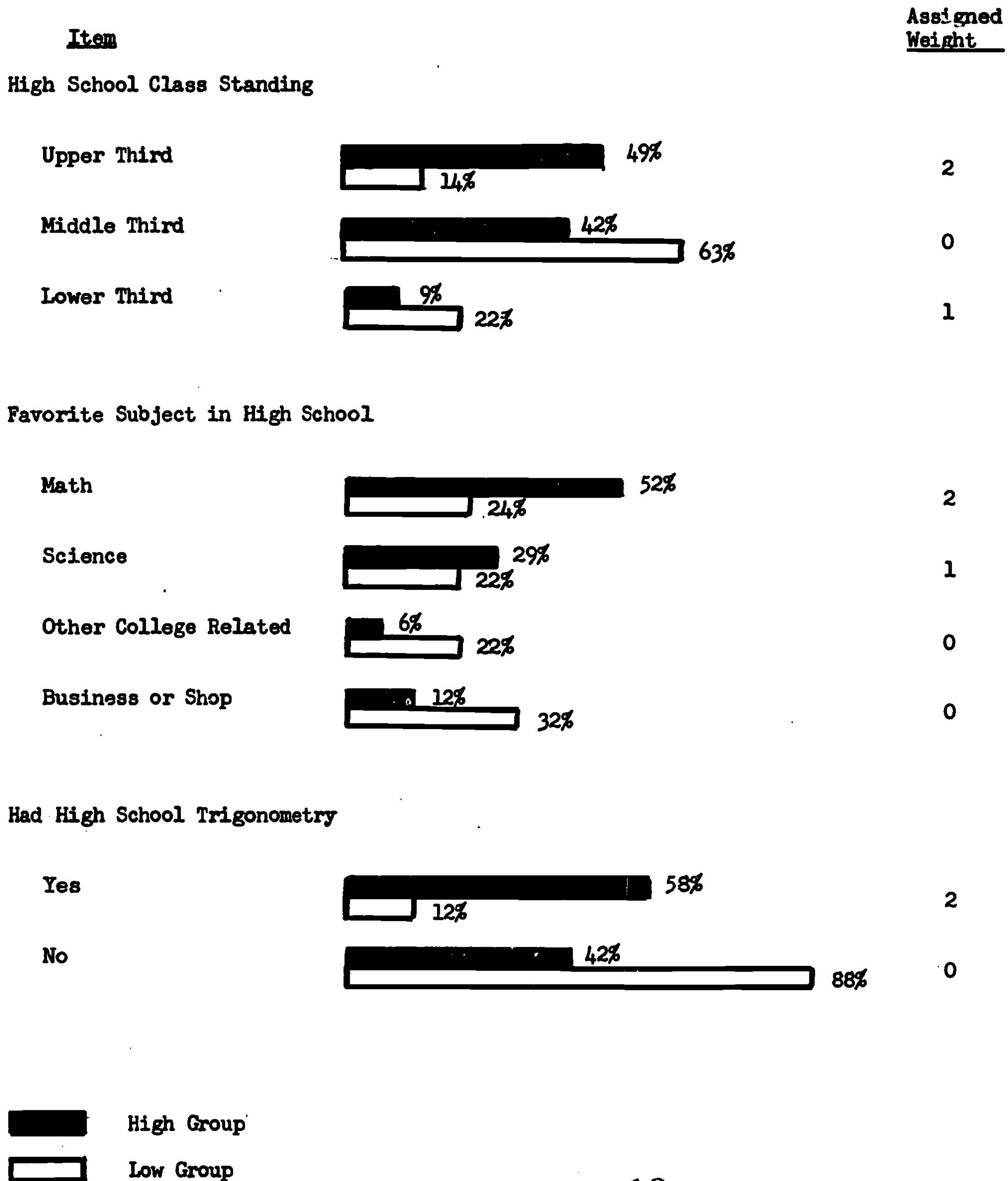
26. Number of related college courses taken (zero, one, two, three or more, not applicable). Deleted because it was being measured by item #11.

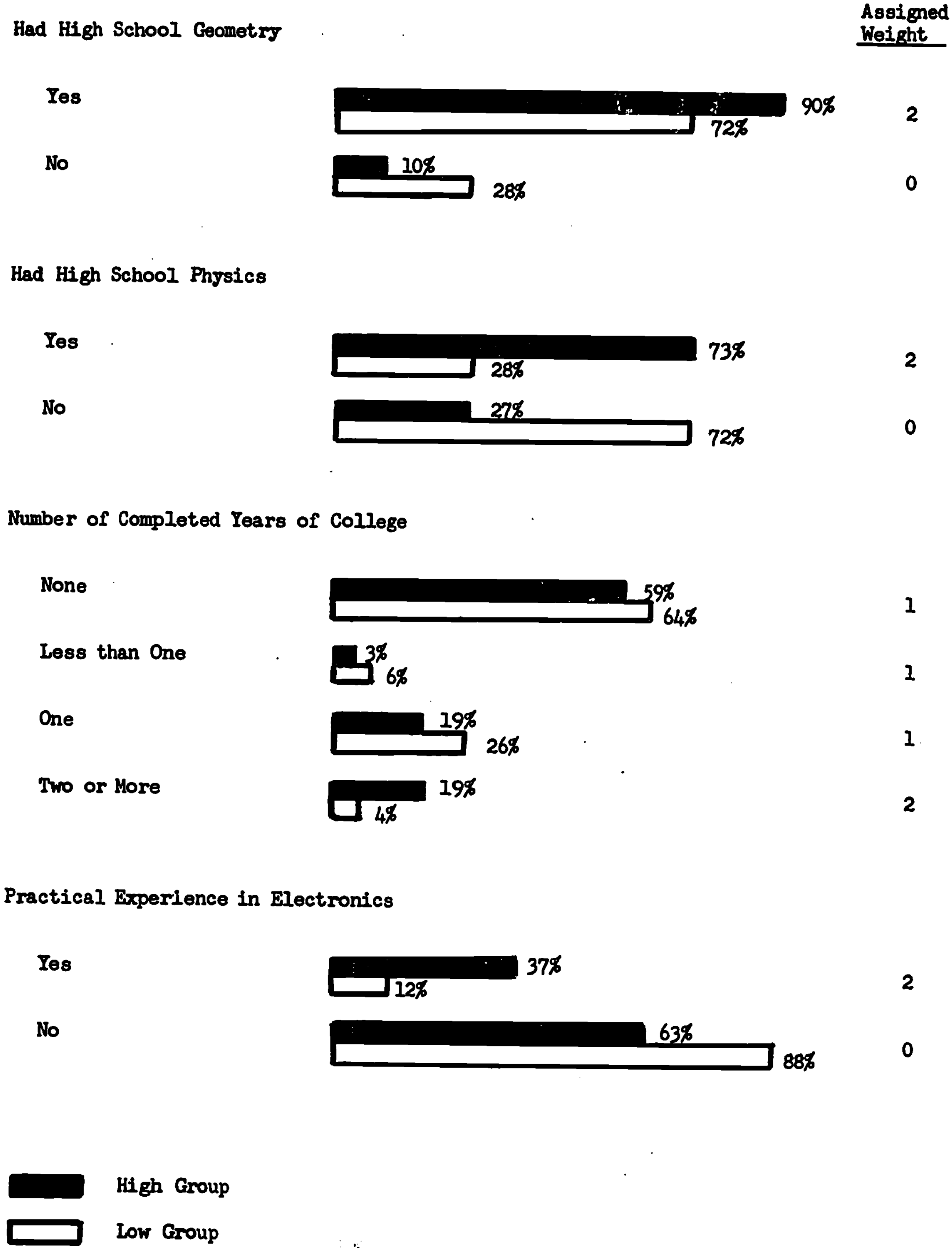
27. Number of years since high school graduation (zero, one, two to four, five or more). Deleted because it measured the same factor as item #10 "Age".

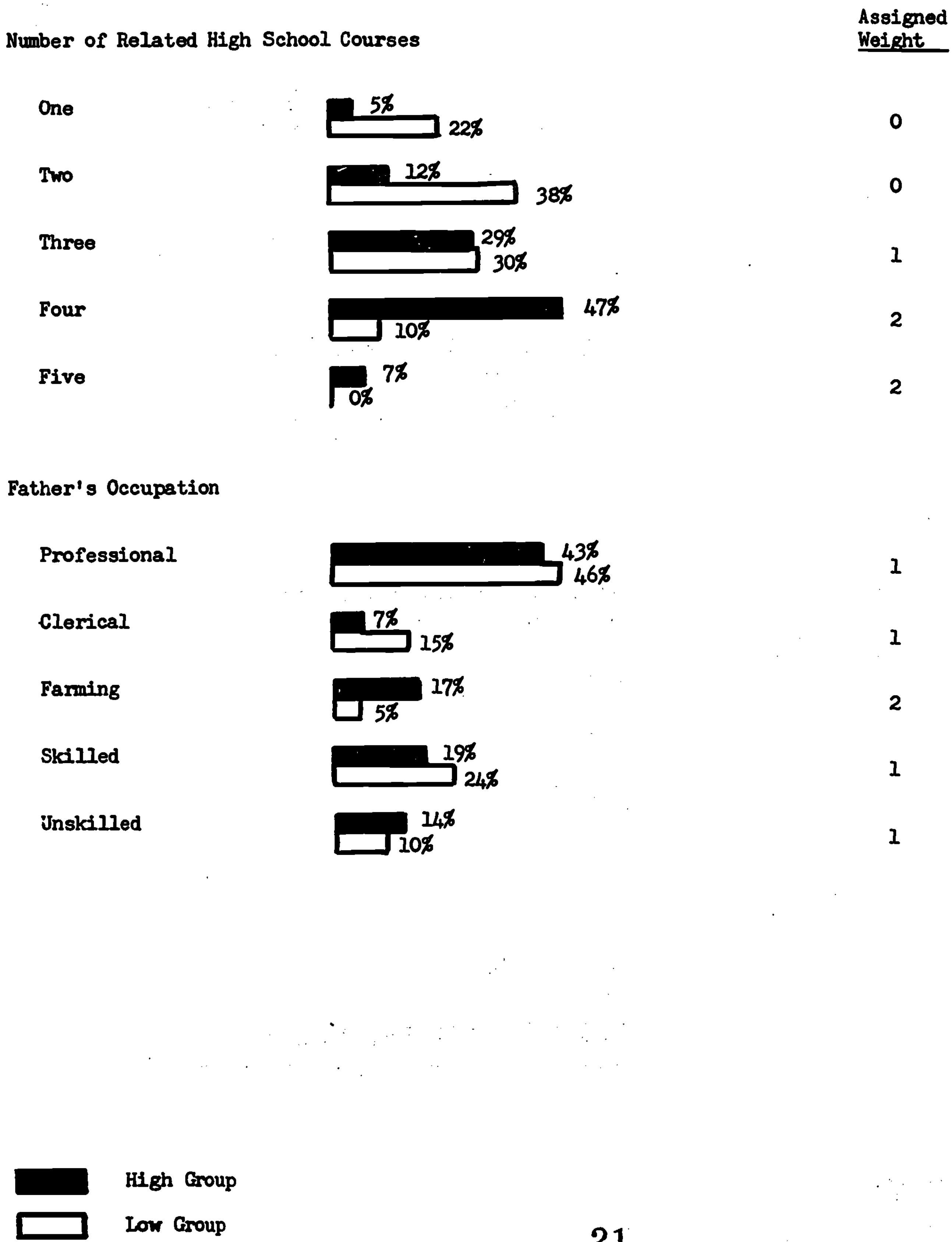
The assigned weights and percentage of good and poor students responding to each category of the differentiating items is graphically portrayed in Table 15.

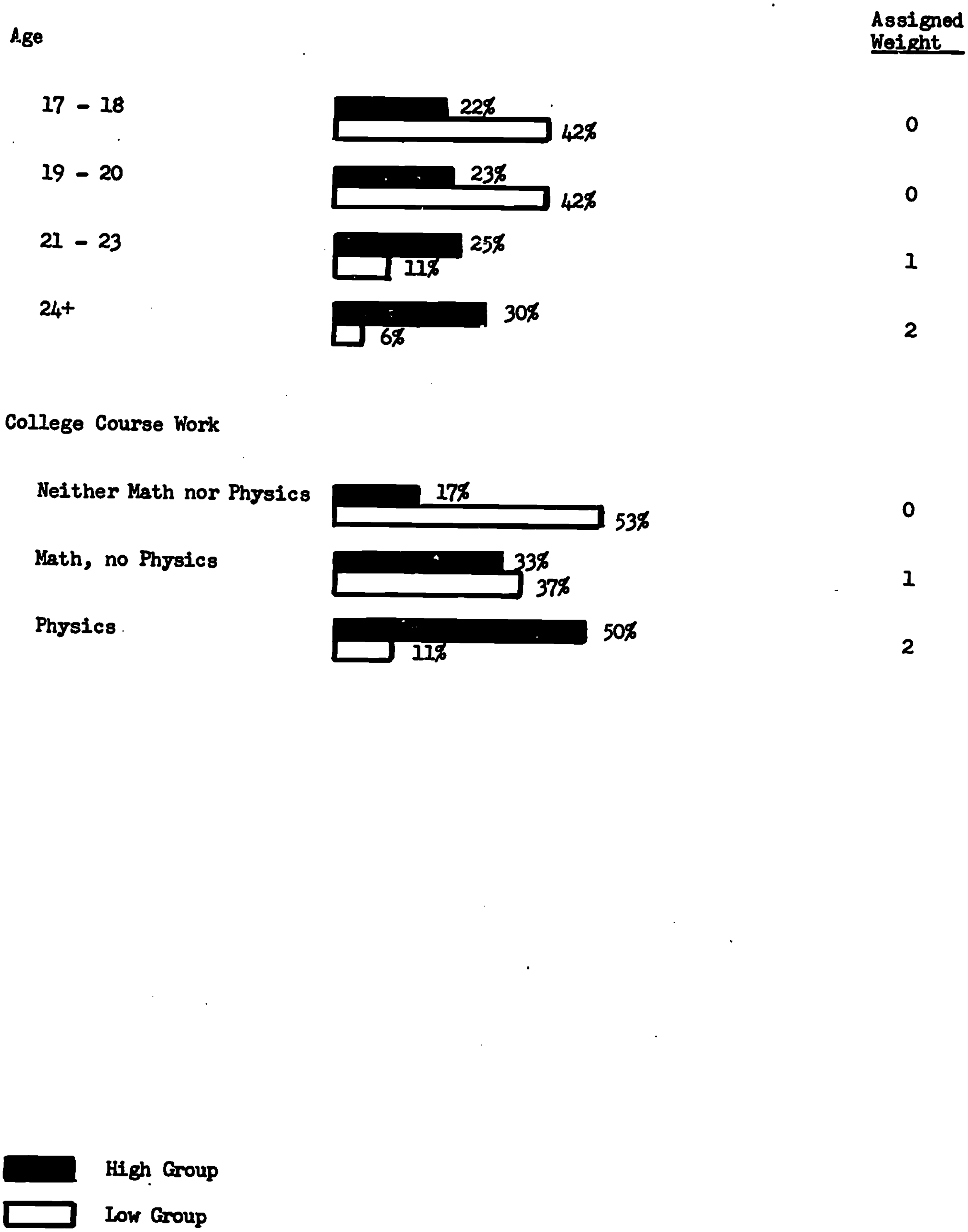
TABLE 15

Percentage of Good and Poor Students Responding to Each WAB Category




Assigned Weight
Had High School Geometry
Yes
90%
72%
2
No
10%
28%
0
Had High School Physics
Yes
73%
28%
2
No
27%
72%
0
Number of Completed Years of College
None
59%
64%
1
Less than One
3%
6%
1
One
19%
26%
1
Two or More
19%
4%
2
Practical Experience in Electronics
Yes
37%
12%
2
No
63%
88%
0
High Group
Low Group


Number of Related High School Courses
Assigned Weight
One
5%
22%
0
Two
12%
38%
0
Three
29%
30%
1
Four
47%
10%
2
Five
7%
0%
2
Father's Occupation
Professional
43%
46%
1
Clerical
7%
15%
1
Farming
17%
5%
2
Skilled
19%
24%
1
Unskilled
14%
10%
1
High Group
Low Group


Age
Assigned Weight
17 - 18
22%
42%
0
19 - 20
23%
42%
0
21 - 23
25%
11%
1
24+
30%
6%
2
College Course Work
Neither Math nor Physics
17%
53%
0
Math, no Physics
33%
37%
1
Physics
50%
11%
2
High Group
Low Group

Validation of WAB:

By adding the assigned weights for the eleven differentiating items it was possible to assign a total score to each individual. For those students who did not answer one of the eleven items on the application blank, the decision was made to use the lowest assigned weight value for that item, which in most cases was zero.

Validation of the WAB was first conducted with the 59 students in the holdout groups using the phi coefficient as a means of evaluation. For the holdout groups a total score of eleven points provided the greatest differentiation between good and poor students ($\phi = .597$, $P/2 < .0005$). As a second test of validity, the phi coefficient was computed for the entire SATB sample even though this represented a partial duplication of students. When applied to the SATB sample, the proportion screened out by a cutting score of eleven was too high. Staying within the q' limits, the best cutting score for the SATB sample was nine, as shown in Table 16.

TABLE 16

Selective Efficiency of WAB-9 for SATB Sample

	Nonqualifying Test Scores	Qualifying Test Scores	Total
Good Students	29	88	117
Poor Students	45	14	59
Total	74	102	176

Corrected $\phi_2 = .480$ $P/2 < .0005$
Corrected $\chi^2_y = 40.6$

It is recommended that this latter cutting score be used because it is based on a larger, more heterogeneous sample than the holdout groups and therefore has wider applicability.

As a second measure of validity, a Pearson product-moment correlation was computed between the WAB and unit exams (the criterion) for the SATB sample. The resulting correlation of .541 was significant at the .01 level.

Cross-Validation of the WAB:

Using the cross-validation sample it was possible to cross-validate the WAB in a manner similar to standard SATB procedures. The computed correlation between the WAB and unit exams was .508 (significant at the .01 level) compared with the .541 found with the original sample. Table 17 shows the four-fold division and phi coefficient for the cutting score of nine.

TABLE 17

Selective Efficiency of WAB-9 for Cross-Validation Sample

	Nonqualifying Test Scores	Qualifying Test Scores	Total
Good Students	33	90	123
Poor Students	31	13	44
Total	64	103	167

Corrected $\phi = .381$
Corrected $\chi^2_y = 24.3$ $P/2 < .0005$

This loss in predictive validity is greater than for most of the research in this study; however, the value of the cross-validated phi coefficient is still high. A comparison between the cross-validated phi coefficients of the WAB and the SATB shows the WAB alone to be almost as effective as the SATB in selecting qualified trainees.

Further Research with the WAB:

Results from the development and validation of the weighted application blank were sufficiently encouraging that various methods of combining the WAB with the SATB in an attempt to improve prediction were tried. Three methods were considered: (1) selecting the most efficient level of the WAB to be combined with the SATB as a fourth cutting score, (2) constructing a double-entry expectancy table[5], and (3) computing a new multiple regression equation including the WAB data.

In order to meaningfully combine the WAB with the SATB using the multiple cut-off procedure, it was first necessary to find the WAB cutting score which would work most efficiently with the SATB norms. Although a cutting score of nine provided the greatest differentiation when the WAB was considered alone, this cutting score eliminated too large a proportion of the sample when included with the SATB. Instead, a cutting score of seven provided the most efficient combination with the SATB for the original study sample. The four-fold division and phi coefficient is depicted in Table 18.

[5]Wesman, A. G., Double-Entry Expectancy Tables, Test Service Bulletin of the Psychological Corporation, May, 1966, No. 56.

TABLE 18

Selective Efficiency of the SATB and WAB-7 for Original Sample

	Nonqualifying Test Scores	Qualifying Test Scores	Total
Good Students	30	87	117
Poor Students	44	15	59
Total	74	102	176

Corrected $\phi = .456$ $P/2 < .0005$
Corrected $\chi^2_y = 36.6$

On the cross-validation sample, this combination of WAB and SATB held up very well, as is demonstrated in Table 19.

TABLE 19

Selective Efficiency of the SATB and WAB-7 for Cross-Validation Sample

	Nonqualifying Test Scores	Qualifying Test Scores	Total
Good Students	32	91	123
Poor Students	34	10	44
Total	66	101	167

Corrected $\phi_2 = .448$ $P/2 < .0005$
Corrected $\chi^2_y = 33.5$

In order to most fully utilize the information contained in the WAB, two other methods of combining it with the SATB were tried, a double-entry expectancy table and a second multiple regression equation.

The double-entry expectancy table was devised as a means for combining two criterion predictors, in this case the WAB and the SATB, to facilitate decision making. Some of the questions answered by such a table are: What is the likelihood of success for a person scoring high on both predictors? What is the probability of success for a person low on both predictors? What are the chances for a person scoring high on one predictor and low on the other? And, which predictor contributes more toward success?

Steps involved in constructing an expectancy table are: deciding where to dichotomize each predictor, counting the number of individuals falling into each cell, and computing the desired percentages for the predictions of success. Since previous research had established the best point of dichotomization for both the WAB and the SATB, it was a very quick procedure to set up the expectancy table. In order to maximize the predictive ability of the table, both the SATB sample and the cross-validation sample were combined for computing the frequencies and percentages. Preliminary tables had already shown very close agreement between the two groups. Table 20 demonstrates how the double-entry expectancy table was constructed and also gives the number of individuals falling in each cell.

TABLE 20

Double-Entry Expectancy Table for Combined Samples

N = 343

	Fail SATB Norms	Pass SATB Norms
High WAB	High Criterion Group N = 22	High Criterion Group N = 156
	Low Criterion Group N = 13	Low Criterion Group N = 14
Low WAB	High Criterion Group N = 17	High Criterion Group N = 45
	Low Criterion Group N = 46	Low Criterion Group N = 30

These frequencies are converted into percentages of success in Table 21.

TABLE 21

Double-Entry Expectancy Table - Percentages of Success

	Fail SATB Norms	Pass SATB Norms
High WAB	63%	92%
Low WAB	27%	60%

Conclusions which may be made from this table are as follows: (1) A person scoring high on both predictors has an excellent chance of succeeding in the training. (2) A person scoring low on both has a poor chance of completing the training. (3) Persons scoring high on one measure but low on the other have approximately a 60% chance for success. (4) A person scoring high on the WAB but failing the SATB has a very slight advantage over a person in the reverse situation, but essentially the two predictors contribute equally toward success.

The final method attempted for combining the SATB with the WAB was to develop a new multiple regression equation incorporating the WAB data. For clarity in the following discussion and tables, this new equation will be referred to as "multiple regression #2". As before, the regression equation was derived and validated on the SATB sample and then cross-validated with the Phase II sample. Applying the Wherry-Doolittle Test Selection Method[6] with factors G, V, N, S, Q, and WAB, the following multiple regression equation was derived:

$$Y' = 1.877X_{WAB} + .252X_G + .168X_S + 1.745$$

where

Y' = Predicted unit examination average

X_{WAB} = Score for weighted application blank

X_G = Score for aptitude G

X_S = Score for aptitude S

The predicted multiple correlation between these factors and the criterion was .593 and the actual attained correlation on the SATB sample was .597. A second means used to evaluate the validity of equation #2 was to determine the best cutting score and compute the phi coefficient. Once again the best cutting score proved to be 67. Table 22 presents the phi coefficient for multiple regression #2.

[6]Garrett, Op. Cit., pp. 426-440.

TABLE 22

Selective Efficiency of Multiple Regression Equation #2 for SATB Sample

	Nonqualifying Test Scores	Qualifying Test Scores	Total
Good Students	26	91	117
Poor Students	46	13	59
Total	72	104	176

Corrected $\phi_2 = .523$ $P/2 < .0005$
Corrected $\chi^2_y = 48.2$

The next step was to cross-validate these results by calculating the correlation between the prediction equation and the criterion and by calculating the phi coefficient on the Phase II sample. The obtained correlation was .603, a value that is even slightly higher than the .597 computed on the original sample. Table 23 gives the computed phi coefficient.

TABLE 23

Selective Efficiency of Multiple Regression #2 for Cross-Validation Sample

	Nonqualifying Test Scores	Qualifying Test Scores	Total
Good Students	28	95	123
Poor Students	33	11	44
Total	61	106	167

Corrected $\phi = .464$ $P/2 < .0005$
Corrected $\chi^2_y = 35.9$

As a means of summary, Table 24 presents the results of the two techniques used to compute the validity of multiple regression #2 and may be compared with Table 9 showing the same data for the first equation.

TABLE 24

Summary of the Validity of the Multiple Regression Equation #2

Obtained correlation between equation and criterion for SATB sample	r = .597
Obtained correlation between equation and criterion for cross-validation sample	r = .603
Phi coefficient for SATB sample	∅ = .523
Phi coefficient for cross-validation sample	∅ = .464

Because there is general agreement that many factors other than ability contribute to job satisfaction and success, particularly in the higher level occupations, it is felt that the results of the data analysis conducted on the WAB justify giving serious consideration to investigating ways of incorporating this information with the SATB norms.

SUMMARY

In a study of this magnitude, involving several new methods and many statistical analyses, it becomes increasingly difficult to compare the results of one part with another part. It was for this reason, that whenever possible, the phi coefficient was used as the method of evaluation and comparison. The following summary table, presenting all of the phi coefficients discussed in this technical report, is designed as an aid to help clarify and snythesize the obtained results.

TABLE 25

Summary Table of Phi Coefficients

Statistical Analysis	Phi Coefficient
SATB norms for:	
validation sample	.416
cross-validation sample	.397
Multiple regression #1 for:	
validation sample	.393
cross-validation sample	.371
Multiple regression #1 and SATB combined for:	
validation sample	.426
cross-validation sample	.375
Tiffin-18 and California-19 for:	
validation sample	.407
cross-validation sample	.262
WAB-11 for holdout groups	.597
WAB-9 for:	
validation sample	.480
cross-validation sample	.381
SATB and WAB-7 for:	
validation sample	.456
cross-validation sample	.448
Multiple regression #2 for:	
validation sample	.523
cross-validation sample	.464

The Minnesota agency feels that the concept of a band group, as applied in this study, is a valid idea deserving further consideration and experimentation. Too often in SATB studies there is no true point of dichotomization between the high and low criterion groups, resulting in the choice of some arbitrary point. Rather than drawing one line in such cases, it might be more accurate to consider a band width between the two criterion groups.

Since the sample size for most test development studies is rather small, it would usually be necessary to have a narrower band width than the one employed for this group of computer technology trainees. For the norm development and validation in this study, 34% of the total usable sample was placed in the band group, and for the cross-validation study the band group comprised 28% of the total sample. In retrospect it is believed that a narrower band width could also have been employed effectively, thereby including a greater portion of the total sample in the SATB development. Some experimentation in this area was tried by the Minnesota agency.

Taking the SATB norms and considering the original sample, the cutting score between the high criterion group and the band group was first lowered from 76 to 70 points. In effect this lowered the percent of students in the band group to 19% and raised the number of students in the high criterion group. The resulting corrected phi coefficient value was .351 (significant at the .0005 level). Then the cutting score was lowered to 65 points, which changed the band group to 7% of the total usable sample. The corresponding phi coefficient was .319 (also significant at the .0005 level). As would be expected, both of these phi coefficients are lower than the value of .416 obtained using the original procedure. In conclusion, when there is no clear-cut line of demarcation between the good and poor group, the use of a small band group may result in obtaining more valid predictors due to the minimizing of misclassification errors. For USES studies a band group not to exceed 20% of the sample would seem to be a reasonable rule of thumb.

RECOMMENDATIONS

Based on the research now completed on the training phase of this study, the following recommendations are made:

1. That the aptitude cutting scores of G-110, N-95, and S-100 be retained as norms for all computer technology training courses similar to the one presented in this study.

2. That further research be conducted with other studies on the value of combining multiple regression with multiple cut-off.

3. That consideration be given to including other predictors of success, such as a weighted application blank, with the SATB.

A-P-P-E-N-D-E-X

Tables Used in the Development of Multiple Regression Equation #2[7]

Table 1

Intercorrelations of WAB, GATB Aptitudes G, V, N, S, Q, and the Criterion

	G	V	N	S	Q	WAB
C	.458	.346	.390	.322	.166	.541
G		.795	.783	.613	.454	.464
V			.518	.327	.314	.307
N				.271	.503	.481
S					.261	.190
Q						.202

[7]Garrett, H. E., Statistics in Psychology and Education, (New York: David McKay Co., 1964), pp. 426-440.

APPLICATION FORM FOR COMPUTER TECHNOLOGY TRAINEES 828.XX

1. Age

    ____17-18

    ____19-20

    ____21-23

    ____24+

2. Father's occupation

    ____professional

    ____clerical

    ____farming

    ____skilled

    ____unskilled

3. High school class standing

    ____upper third

    ____middle third

    ____lower third

4. Favorite subject in high school

    ____math

    ____science

    ____other college related course

    ____business or shop course

5. Which of the following courses did you take in high school

    trignometry ____yes ____no

    geometry ____yes ____no

    physics ____yes ____no

6. How many of the following courses did you study in high school: (1) algebra (2) geometry (3) trigonometry (4) physics (5) electricity or radio

    ____one

    ____two

    ____three

    ____four

    ____five

7. Number of completed years of college

    ____none

    ____less than one

    ____one

    ____two or more

8. College course work

    ____studied neither math nor physics

    ____studied math but not physics

    ____studied physics

9. Have you had practical work experience in electronics

    ____yes

    ____no

GPO 870-625

U.S. DEPARTMENT OF LABOR
MANPOWER ADMINISTRATION
BUREAU OF EMPLOYMENT SECURITY

OFFICIAL BUSINESS

POSTAGE AND FEES PAID
U.S. Department of Labor

THIRD CLASS MAIL