1-1-1965

Psychological tests in the selection of hair accessory carders

Craig B. Edwards
University of Nebraska at Omaha

PSYCHOLOGICAL TESTS IN THE SELECTION OF HAIR ACCESSORY CARDERS

A Thesis

Presented to the

Department of Psychology

of the

Faculty of the College of Graduate Studies

University of Omaha

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Craig B. Edwards

January, 1965

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

DEDICATION

I dedicate this thesis to my wife, Merry, without whom it never would have been started; and to my father for his deep belief in the value of education. I would also like to thank Dr. William Jaynes for his guidance in preparing this paper.

C. B. E.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

THE PROBLEM

Since the beginning of time, men have been trying to decrease the cost of production by increasing the output of the individual. Slavery, paternalism, motivation, and psychological tests are just a few of the methods devised to increase the output of the work group. Many kinds of knowledge have been brought to bear on this problem. Studies, explanations, and proposed solutions have come from such varied disciplines as philosophy, economics, religion, sociology, and psychology. Neither individually nor collectively have these contributions provided any final solution for the problem of decreasing the cost of production by increasing the output of the individual.

I. STATEMENT OF THE PROBLEM

The XYZ company employed a large percentage of Omaha's female common labor supply in its manufacturing process. The work these women did consisted mainly of placing hair curlers and barrettes into pasteboard cards.[1]

---

[1]Most employees work in this department at one time or another.

The women were paid on an incentive wage plan. However, under the federal minimum wage law, all carders were paid a specified minimum rate[2] whether their actual production was in excess of this minimum or not.

The production of the average carder did not exceed the necessary production to earn the minimum wage for approximately six weeks.[3] The difference between the actual production of the first six weeks of employment and the required minimum wage cost the company about twenty-eight dollars per new employee, of which there were at least five hundred per year.

In order to decrease the production costs of the product, it was necessary to reduce the cost of the six-week learning period by hiring workers who would produce at a high rate over an extended period of time and/or hiring applicants who would raise their production to the minimum more quickly than those hired in the past. Better pre-employment selection would be one method through which this aim could be accomplished. The goal of this study was to develop such a pre-employment selection program using the Wonderlic Personnel test, the Purdue Pegboard,

---

[2]The minimum wage was $1.00 per hour when this study was originally undertaken. On September 4, 1961, the minimum wage became $1.15.

[3]See Figure 1, page 16.

and the Thurstone Temperament Schedule.

II. IMPORTANCE OF THE STUDY

Capital is the beginning of an industrial organization. It buys physical facilities, tools, raw materials, and human energy. All of these interact to produce a finished product. If, however, the price the consumer is willing to pay for the product is less than the cost to produce and market the goods, the production system must become more efficient or the capital investment must be used to make up the deficit. The continued expenditure of initial capital, without new investment or increased profits, must lead to the dissolution of the enterprise. It is therefore necessary that an enterprise, in order to remain in business and make jobs, make profit (the remainder after all factors of production have been fully compensated)[4] its prime concern. In order that a profit be made, a company must strive to obtain a maximum return per dollar invested in the production process, each item which contributes to the cost of the finished goods must be returning as near the maximum as possible.

The American Economic Foundation has published a

---

[4]Erwin Nemmers, Dictionary of Economics (Patterson: Littlefield, Adams and Company, 1959), p. 234.

pamphlet called, How and Why to Prepare the Functional Operating Report, which outlined the six costs of operating a business.[5] The six costs were as follows: (1) income from the customer; (2) cost of tools wearing out; (3) cost of human energy; (4) cost of payments ordered by the government (taxes); (5) cost of using tools; (6) cost of goods and services bought from others. More simply stated: MMW = NR + (HE x T); man's material welfare (MMW) is equal to natural resources (NR) plus human energy (HE) multiplied by tools (T).

Only the cost of human energy was within the scope of this paper. If the cost of human energy could be reduced, the primary aim of the industrial firm (increased profits) could, to some degree, be achieved.

III. HISTORY OF THE PROBLEM

For thousands of years, men have thought that the use of slave labor was the most efficient and inexpensive method of production. However, the use of slaves on southern plantations just before the Civil War was down considerably from what it had been a decade or two before. A survey of southern plantation owners taken just before the

---

[5] How and Why to Prepare the Functional Operating Report (New York: The American Economic Foundation, 1958), p.6.

Civil War indicated that the reason for the decrease in the use of slaves was the fact that free men produced much more for wages than the slave for room and board.[6] Slavery, as a system of cheap labor, not only did not solve the problem of the cost of human energy but increased it.

During the middle ages, the guild system flourished and then died and was replaced by the putting-out system or "domestic system" as it is more usually called. This system existed side by side with both the guild system and the factory system. In the putting-out system, the worker was not bound to a master; he usually owned his own tools; and he, and his whole family, worked in their own home. His raw materials were given him by the merchant capitalist. He was paid for his labor, but the product belonged to the merchant. Schnieder describes the decline of this system in this way:[7]

> As time went on, the formal relations of production became increasingly capitalistic in nature. From the point of view of the merchant, the putting-out system was hopelessly inefficient. It was extremely difficult to supervise the labor of scattered workers, and there was great loss through waste and embezzlement. Furthermore, the labor supply was uncertain and shifting.

---

[6]Robert LeFavre, of the Freedom School, in a lecture, August, 1961.

[7]Eugene V. Schnieder, Industrial Sociology (New York: McGraw-Hill Book Company, 1957), p. 38.

In addition to the carelessness and thievery of the workers, the process by which the goods were produced under this system involved moving semi-completed goods from one place to another. This, of course, involved a great deal of the workers' time and energy. The amount of goods produced was therefore meager and the cost was staggering. As demands for manufactured goods increased, it was discovered that production could never keep up with the demand. Thus a new system, the factory system, was a natural outcome.

CHAPTER II

CURRENT LITERATURE AND DEFINITION OF TERMS

The nature and significance of individual differences among industrial employees have been of constant interest to the psychologist. The journals are filled with papers concerning this vital area of study. Among the growing number of articles, being published each year, are many concerning the testing and selection of factory workers. Several of these articles pertained either to the general type of employee that this paper dealt with or to the type of tests used in this study.

I. CURRENT LITERATURE

The question concerning the value of intelligence tests in predicting production of workers is still very much up in the air. Some researchers have found that intelligence tests were useful in predictive studies; others found that they showed very little promise. Ghiselli and Brown stated, "Intelligence test scores do predict proficiency in production on some jobs. . . .For . . . . assemblers, and gross manual workers the validity of intelligence tests is fairly good."[1] Ghiselli and Brown

---

[1]Edwin Ghiselli and Clarence Brown, Personnel and

claim that the relation between intelligence and the proficiency of assemblers was about .22 as indicated by a Pearson product-moment correlation coefficient.[2] Tiffin and Greenly reported a negative relation between routine manipulative assembly and the Otis Self-Administrating Test of Mental Ability.[3] Cuomo and Meyer reported, in the validity exchange, a product-moment correlation of .17 between the Wonderlic test scores and the speed of upgrading of ninety floor assemblers.[4] Another study of fifty light-bulb assemblers reported a product-moment correlation of .48 between the Wonderlic test score and supervisory ratings.[5] Many studies, using the various tests of verbal intelligence, have a curved regression. Guilford stated:[6]

Industry Psychology (New York: McGraw-Hill Book Company, 1955), p. 237.

[2]Ghiselli and Brown, op. cit., p. 234.

[3]Joseph Tiffin, Industrial Psychology 3rd Ed. (Englewood Cliffs: Prentice-Hall, Incorporated, 1956), p. 110.

[4]Sylvia Cuomo and Herbert H. Meyer, "Validity Information Exchange, #7-077:D.O.T. Code 6-78.632 Floor Assembler," Personnel Psychology 8:270 Su' 55.

[5]U. S. Employment Serv. "Validity Information Exchange, #7-093 D.O.T. Code: 7-00.070 Light Bulb Assembler" Personnel Psychology Vol. 7. W. 1954 p. 571.

[6]J. P. Guilford, Fundamental Statistics in Psychology and Education (New York: McGraw-Hill Book Company, 1956), pp. 387-388.

> There has been a common conclusion in the industrial psychology literature that individuals of high intelligence are likely to do less well at highly routinized, ... the fact of curved regression is undeniable and should be recognized in selection.

In reviewing the literature Ghiselli and Brown found thirty studies of the relation of finger desterity tests and job proficiency for assemblers and bench workers.[7] Of the thirty studies, four had correlation coefficients of -.05 to .09. Twenty-three of the studies had validities ranging from .10 to .49; the remaining three studies were in the .55 to .64 and .85 to .89 area. The median of the coefficients was in the area of .25 to .29.

A study by L. V. Surgent yielded a validity coefficient of .64 between the Purdue Pegboard Assembly Subtest and ratings of supervisors and a correlation of .22 with productivity as indicated by earnings.[8] Tiffin and Greenly reported of electrical fixture assemblers and radio assemblers a correlation of .33 with general efficiency as indicated by merit ratings.[9]

---

[7]Ghiselli and Brown, op. cit., p. 220.

[8]L. V. Surgent, "The Use of Aptitude Tests in the Selection of Radio Tube Mounters," Psychological Monograph, 1947, 61, No. 2, 1-40.

[9]J. Tiffin and R. J. Greenly, "Employee Selection Tests for Electrical Fixture Assemblers and Radio Assemblers," Journal of Applied Psychology, 1939, 23, 240-263.

There seemed to be a void in the literature concerning the use of personality inventories to select factory workers in general and assemblers in particular. Either no studies in this area were undertaken or, if they were undertaken, they had not been published, which might indicate negative results. However, in other job areas, the personality inventory has been used with some success. Ghiselli and Brown showed average correlations of .30, .24, and .27 between personality test scores and the proficiency of processing workers, complex machine operators, and for all trades and crafts.[10] The validity of personality inventories in predicting job proficiency for sales and protective workers was not as good as for the trades and crafts area but was promising.

In reviewing the literature, it was immediately evident that there was still much to be learned about the different tests and their predictive value. This was particularly true in the area of personality tests.

II. DEFINITION OF TERMS

In order to avoid misunderstanding, terms which were

---

[10] Ghiselli and Brown, op. cit., pp. 228-233.

unusual or used in an unusual manner in this paper were defined below.

Validity

Concurrent. Concurrent validity is an empirical check on the agreement between the criterion and the test(s) when both the criterion and the predictive information are obtained at the same or nearly the same time.

Predictive. Predictive validity is an empirical check on the agreement between the criterion and the predictor when the criterion is obtained at some time subsequent to the predictor.

Reliability

Split-half. Split-half reliability is obtained by dividing a test into two equal halves for each subject, then correlating the two halves with each other. Split-half reliability provides a measure of equivalence, or adequacy of item sampling.

Retest. The test-retest reliability is obtained by repetition of the identical test on a second occasion. This type of reliability provides a measure of temporal stability.

Compound Probability

Given two independent observations of a single hypothesis, the two observations yielding probabilities p_1

and p_2 for the hypothesis; then the probability statement based upon the two observations taken together is the compound probability. This compound probability is given by the probability of a Chi Square with four degrees of freedom where $X^2 = -2 \log_e p_1 p_2$.[11]

---

[11] Paul C. Baker, "Combining Tests of Significance in Cross Validation, "Educational and Psychological Measurement, 1952, 12, 300-306.

CHAPTER III

THE CRITERIA AND THE TESTS

It is impossible to say that any one particular criterion is always best. The choice of the criterion is dependent upon two things: (1) the object of the study and (2) the availability of a reliable and valid criterion. If either the validity or the reliability of the criterion is in doubt, the results of a study are of little value.

I. THE CRITERIA

Quality of work, tenure, absenteeism, accidents, merit rating and production are some of the most widely used criteria in present day industrial research. In the case of jobs which are paid on an incentive plan, the production figures which are necessarily available are the most objective and readily available measure of productivity.

The Prime Criterion

The productivity of the worker was of prime importance to the XYZ company. It was felt by the management that an upgrading of their employees was of vital importance to the company's continued growth. The natural focus of attention was upon the source of their labor supply, the employment process. As the pre-employment testing of applicants was done rather haphazardly, it was decided that an

attempt should be made to institute a systematic pre-employment testing program.

At this point it was necessary to determine two things: (1) the criterion or criteria to be used and (2) the group to be studied. As most of the hourly workers were on some form of incentive wage plan, the most efficacious and objective measure of the individual workers productivity was their incentive wage which bore a direct relationship to the number of pieces produced. This was selected as the general criterion of the study. It must be understood that the incentive wage used in this study as a criterion was computed directly from the production of the worker and was not in any way effected by the Federal Minimum Wage Law.

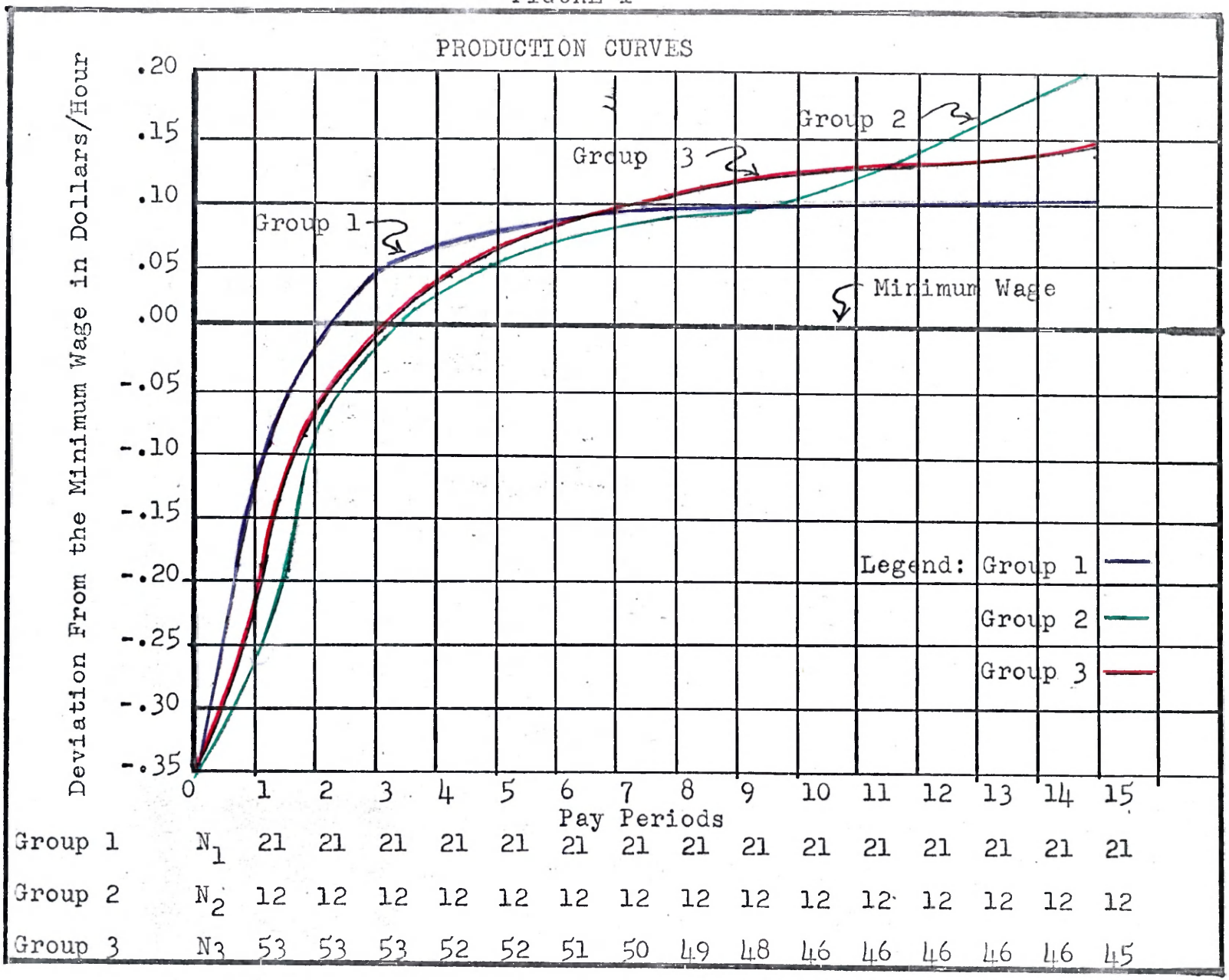
As one of the goals of this study was the prediction of long-term success, some measure of long-term success had to be derived from the general criterion. The mean wage in terms of deviations from the minimum wage covering the first year of employment seemed like a good criterion. Upon inspecting the employees in the department, it was found that only ten carders had been in the department continuously for a year or more. This was too small a sample upon which to base a study. It was therefore necessary to find a time period in which the number of carders was large enough for study and the mean deviant wage was still indicative of

long-term success.

The production records of all employees, both past and present, whose first working assignment was in the carding department were converted to deviant wages. In order to prevent criterion distortion, any carder who was transferred to some other classification for more than two pay periods was dropped from the study. The mean deviant wage for each pay period was then plotted to determine when the carders' production leveled out. Group III in figure 1 shows the results of this portion of the study. The vertical axis represents deviations from the minimum in cents; the horizontal axis represents the time in the job classification in pay periods. Group I consisted only of the carders in the concurrent validity study. Group II consisted of those carders in both the predictive validity and cross validation studies. Group I is a subgroup of Group III.

Each pay period was equal to two weeks. At the end of the first pay period, the average worker's production, using Group III as an estimate of the universe being studied, was about nineteen cents below the minimum wage. By the fourth pay period, the mean deviant wage had risen $0.248 to $0.062 per hour above the minimum wage. Those who remained a full year in the job classification studied earned about $0.16 per hour above the minimum wage.

FIGURE 1



Group 1	N_1	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
Group 2	N_2	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
Group 3	N_3	53	53	53	52	52	51	50	49	48	46	46	46	46	46	45

Correlations between each pay period and the one immediately following yielded test-retest reliability coefficients which gave some indication of the stability of the workers production from pay period to pay period. Those pay periods which showed the most stability were selected as the criterion but due to the rapid decrease in sample size after the first fifteen pay periods, only the first fifteen pay periods were considered.

The test-retest reliabilities for the first nine pay periods ranged from .135 to .532 with a median of .447. The test-retest reliabilities for the remaining pay periods ranged from .616 to .652 with a median of .631. Therefore, the tenth through fourteenth pay periods inclusive were designated the prime criterion.

The Secondary Criterion

In addition to the prime criterion already discussed, a secondary criterion was established. The average carder took three pay periods before her production was equal to the minimum wage, and it was almost eight pay periods before two thirds of the carders were producing this amount. Therefore, it would be advantageous to hire carders who could raise their production to the break even point more quickly.

The carders were placed into two groups depending upon whether their production during the first six pay periods

was above the mean production for the group. The placement of a carder into one of the two groups constituted the secondary criterion. Although some information was lost by dichotomizing a continuous variable, the loss was more than compensated by the ease of computation and interpretation to the personnel director of the XYZ company. Due to this procedure, it was necessary to use biserial correlations rather than product-moment correlations in the validity computations.

II. TESTS USED IN THE STUDY

Three tests were used in this study: a test of intelligence, a test of finger dexterity, and a temperament test. The intelligence test and the temperament test had previously been used sporadically for selection and placement.

The Wonderlic Personnel Test - Form F

The Wonderlic is a timed, paper and pencil test of intelligence.[1] The actual testing time is twelve minutes and scoring takes about a minute. The test is almost completely self-administering. It can be given, scored, and interpreted by a clerk with a little instruction. The test

---

[1] E. F. Wonderlic, Manual of the Wonderlic Personnel Test (Northfield: Wonderlic, 1961), pp. 3-7.

is available in nine forms (one form is available only for employment agencies). The form used in this study, Form F, is adapted from the Otis Self-Administering Tests of Mental Ability, Higher Form. Six different normative groups are included in the manual; the "total" group containing 53,864 subjects.

The test-retest reliability of the test ranged from .82 to .93, indicating good reliability. Split-half reliabilities of .88 to .94 also indicated good reliability.

Anastasi, in her book, Psychological Testing, said, "Despite its brevity, it [the Wonderlic Personnel Test] correlates highly with the original Otis test correlations from .81 to .87 having been obtained from various groups."[2] Thus the Wonderlic Personnel Test seemed to have concurrent validity as far as longer tests of the same type were concerned.

The Purdue Pegboard

The Purdue Pegboard is a performance test of manual dexterity which, although usually given individually, can be given in small groups.[3] The test consists of five different

---

[2]Anne Anastasi, Psychological Testing (New York: MacMillian Company, 1956), p. 229.

[3]L. L. Thurstone, Examiner Manual for the Purdue Pegboard (Chicago: Science Research Associates, Incorporated,

subtests. This test provides a measure of two types of activity, one requiring gross movements of the hands, fingers, and arms, and the other involving "tip of the finger" dexterity needed in small assembly work. The first subtest (R) requires the subject to place pins, one at a time with the right hand, into small holes lined vertically down the board. This operation is repeated using the left hand (L) and both hands together (B). The fourth subtest is simply the total of the scores of the preceding three subtests (R+L+B). The final subtest, Assembly, requires the use of pins, washers, and collars. A pin is placed into the top hole in the right hand vertical row by the subject's right hand; at the same time, the left hand picks up a washer and places it over the pin; the right hand then picks up a collar and places it over the pin; the left hand then picks up another washer and places it over the pin, completing the assembly. This process is repeated until the time limit of one minute is reached. The score is the number of pieces placed properly within the time allowed.

The manual of the Purdue Pegboard gave data concerning reliability, validity, and normative groups for both one and three trials. The three trial method was used in the present

---

1948), pp. 1-8.

study. The one trial test-retest reliabilities ranged from .60 for the Pegboard-L to .91 of the Pegboard-Assembly, with a median of .86. The one three trial test-retest reliability reported was .64 for the Pegboard-Assembly. The validity coefficients based on one trial results ranged from .07 to .76 with a median of .36. Two studies reported in the manual, using production as a criterion, had validity coefficients of .76 while a third study using the same type of criterion yielded a validity coefficient of .15 with one trial scores. Five different normative groups were included in the manual: 481 college men; 392 college women; 1958 veterans; 865 male industrial applicants; and 4138 female industrial applicants. This test had only one form.

Thurstone Temperament Schedule

During the years 1934 to 1943, J. P. Guilford identified thirteen factors of personality.[4] In 1951, L. L. Thurstone did a second-order factor analysis with the correlations reported by Guilford in his original study. This schedule is based upon the seven second-order factors found by Thurstone.

The Thurstone Temperament Schedule is a pencil-and-paper test consisting of 140 items each of which has three

---

[4] Anastasi, op. cit., p. 537.

possible alternative answers: Yes, ?, No. The test can either be hand scored using the built-in carbon answer pad or machine scored using IBM equipment. The test can be administered to an individual or a group and has no time limit. The factors were described as follows:[5]

> Activity (A). A person scoring high in this area works and moves rapidly. He is restless whenever he has to be quiet. He likes to be "on the go" and tends to hurry. He usually speaks, walks, writes, drives, and works rapidly, even when these activities do not demand speed.
>
> Vigorous (V). A person with a high score in this area participates in physical sports, work requiring the use of his hands and the use of using large muscle groups and great expenditure of energy. This trait is often described as "masculine," but many women and girls will score high in this area.
>
> Impulsive (I). High score in this category indicates a happy-go-lucky, daredevil, carefree, acting-on-the-spur-of-the-moment disposition....The decision to act or change is quick regardless of whether the person moves slowly or rapidly (Active), or enjoys or dislikes strenuous projects (Vigorous).
>
> Dominant (D). People scoring high on this factor think of themselves as leaders, capable of taking the initiative and responsibility. They are not domineering, even though they have leadership ability. They enjoy public speaking, organizing social activities, promoting new projects, and persuading others. They are the ones who would probably take charge of the situation in case of an accident.
>
> Stable (E for Emotional Stability). Persons who have high Stable scores usually are cheerful and have an even disposition. They can relax in a noisy room, and they remain calm in a crisis. They claim that they can disregard distractions while studying. They are

---

[5] Anastasi, op. cit., pp. 1-2.

not irritated if interrupted when concentrating, and they do not fret about daily chores. They are not annoyed by leaving a task unfinished or by having to finish it by a deadline.

Sociable (S). Persons with high scores in this area enjoy the company of others, make friends easily, and are sympathetic, cooperative, and agreeable in their relations with people. Strangers readily tell them about personal problems.

Reflective (R). High scores in this area indicate that a person likes meditative and reflective thinking and enjoys dealing with theoretical rather than practical problems. Self examination is characteristic of reflective persons. These people are usually quiet, work alone, and enjoy work that requires accuracy and fine detail work. They often take on more than they can finish, and they would rather plan a job than carry it out.

The test manual reported both split-half and test-retest reliabilities. The test-retest reliabilities ranged from a low of .61 for the E score to .82 for the D score with a median of .78. The split-half reliabilities corrected for length ranged from .45 for the R score to .86 for the D score with a median of .65.[6]

Only two correlation studies of validity were included in the manual. The first study related the scores on the test to forced choice ratings made by supervisors of retail salesmen. The biserial correlations between the criterion and the test scores extend form a low of .81 for the R score to 1.00 for the I score. The number of subjects was

---

[6]Anastasi, op. cit., p. 6.

not reported.[7]

The second study concerned retail store managers. The manual reported Pearson product-moment correlations between the test scores and a "success" score. The correlations ranged from a low of -.08 to a high of .63 with a median of .41. The correlations between the criterion and the A, D, E and S were significant beyond the .01 level; the correlation with the I score was significant between the .05 and .01 level. The mean sales per month was used as a second criterion in this study. Sixty-one managers were used in this study. The Spearman rank-order correlation technique was used to determine the degree of relationship. The rank-order correlations ranged from .00 for V to .51 for D with a median of .31.[8]

Van Steenberg,[9] in his review of the Thurstone Temperament Schedule for the Fifth Mental Measurements Year Book, said the following about the test's validity possibilities:

> The schedule should prove useful for employment managers and counselors, though it would be desirable to have a number of additional validity studies published . . . . the instrument seems to be well established as an "anchor" for further research on personality.

---

[7] Ibid., p. 10

[8] Anastasi, op. cit., p. 13.

[9] N. J. Van Steenberg, "The Thurstone Temperament Schedule" Fifth Mental Measurements Year Book (Highland Park: Gryphon Press, 1959) p. 207.

CHAPTER IV

THE METHODOLOGY AND RESULTS

Having selected the criterion as indicated in Chapter III, a check of the company records was made to obtain the production records of all those presently working in the carding department. The total number of employees in the department at that time was thirty-four. Some of these women had been employed by the XYZ company only a few weeks while one woman had been employed by the company and worked in this department for over five years.

A date for testing these women was set. Of the thirty-four women who had been in the department when it was originally selected as the experimental group only twenty-six remained for testing. The relation between the test scores and the criteria as indicated by the Pearson product-moment correlation coefficient was used to determine the concurrent validities for the individual tests. These concurrent validity coefficients were then employed as guides for further testing using applicants as subjects in order to establish the predictive validity for each test.

I. THE DEAF SUBGROUP

Within the group tested were four deaf girls. All were able to sign, but only one was able to read lips. There

was a possibility that their abnormality might effect their test scores; thus adding to the difficulty of interpreting results. As the cell frequencies were very small and in many cases fell below the minimum required to use the Chi Square technique with the Yates correction, the Fisher Exact Probabilities Test[1] was used to determine if the four deaf girls differed in their test scores or production from the twenty-two normal girls. Each variable was dichotomized at the mean. The results are shown in Table I. The scores of the two groups were significantly different at the .10 level for the Wonderlic and the Pegboard-R, and at the .05 level on the Pegboard-B and Pegboard-R+L+B. Even though there was some evidence that the Deaf group was more dexterous and had less verbal ability than normal, it was decided to leave them in the experimental group.

II. RESULTS IN CONCURRENT GROUP

Of the twenty-six carders who took the tests, only twenty-one completed the fourteen pay periods required. To establish concurrent validity, a Pearson product-moment correlation between each test and the prime criterion was

---

[1]Sidney Siegel, Nonparametric Statistics for the Behaviorial Sciences (New York: McGraw-Hill Book Company, 1956), pp. 96-104.

TABLE I

COMPARISON OF THE SCORES OF NORMAL AND DEAF CARDERS ON DIFFERENT VARIABLES

Variable	Group	Normal P	Deaf P
Wonderlic	Above Mean	.54	.00*
	Below Mean	.31	.15
Pegboard-R	Above Mean	.31	.15*
	Below Mean	.54	.00
Pegboard-L	Above Mean	.54	.15
	Below Mean	.31	.00
Pegboard-B	Above Mean	.27	.15**
	Below Mean	.58	.00
Pegboard-R+L+B	Above Mean	.20	.15**
	Below Mean	.65	.00
Pegboard-Assembly	Above Mean	.34	.08
	Below Mean	.50	.08
Temperament-A	Above Mean	.42	.115
	Below Mean	.42	.045
Temperament-V	Above Mean	.42	.16
	Below Mean	.42	.00
Temperament-I	Above Mean	.39	.15
	Below Mean	.46	.00
Temperament-D	Above Mean	.39	.11
	Below Mean	.46	.04
Temperament-E	Above Mean	.42	.08
	Below Mean	.42	.08
Temperament-S	Above Mean	.34	.08
	Below Mean	.50	.08
Temperament-R	Above Mean	.39	.08
	Below Mean	.46	.08

*Difference between normal and deaf carder significant at the .10 level.

**Difference between normal and deaf carder significant at the .05 level.

TABLE II

MEDIANS, MEANS, AND STANDARD DEVIATIONS IN THE CONCURRENT GROUP

Variables	Median	Mean	Standard Deviations
Prime criterion	0.072	0.085	.0.108
Secondary criterion	0.028	-0.015	0.084
Wonderlic	14.000	12.333	5.397
Pegboard-R	53.000	54.714	7.099
Pegboard-L	48.000	46.881	4.705
Pegboard-B	41.000	42.310	5.753
Pegboard-R+L+B	139.000	144.643	14.440
Pegboard-Assembly	116.000	120.955	14.261
Temperament-A	10.000	10.452	2.746
Temperament-V	8.000	7.595	3.100
Temperament-I	11.000	10.738	3.379
Temperament-D	6.000	6.119	3.722
Temperament-E	7.000	8.071	3.458
Temperament-S	9.000	9.690	3.512
Temperament-R	7.000	7.881	3.302

TABLE III

INTERCORRELATIONS OF TESTS AND PRIME CRITERION: CONCURRENT GROUP

Variables	2	3	4	5	6	7	8	9	10	11	12	13	14
1	-.40	-.27	-.15	-.41	-.12	.33	.09	.41	.26	.04	.28	.45*	.18
2		.50*	.38	.90**	.36	.07	.22	.14	.08	-.07	-.24	.03	.20
3			.70**	.80**	-.80**	-.28	.14	.09	-.06	.11	-.19	-.07	.53*
4				.67**	.78**	.05	.33	.12	.10	.10	.00	-.09	.52*
5					.84**	.05	.22	.12	.03	-.01	-.16	.00	.39
6						.24	.35	.20	.11	-.09	-.13	.09	.70**
7							.50*	.72**	.50*	.27	.48*	.44*	.26
8								.60**	.53*	.27	.60**	.28	.12
9									.58**	.44*	.68**	.25	.03
10										.33	.50*	.15	.19
11											.58**	.18	.07
12												.11	-.18
13													-.02

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

KEY:

Test	Variable	Test	Variable
Wonderlic	1	Temperament-V	8
Pegboard-R	2	Temperament-I	9
Pegboard-L	3	Temperament-D	10
Pegboard-B	4	Temperament-E	11
Pegboard-R+L+B	5	Temperament-S	12
Pegboard-Assembly	6	Temperament-R	13
Temperament-A	7	Prime criterion	14

computed using these twenty-one sets of scores. The results are shown in Tables II and III.

Concurrent Validity: Prime Criterion

The tests were then combined in a battery by use of the Wherry-Doolittle Test Selection method.[2] The tests which were selected as part of this battery were to be given to applicants who were hired as future carders in order to cross validate the battery and conduct a predictive validity study. Due to the additive nature of the Pegboard R+L+B and the inclusion of the individual subtests R, L, and B were in the correlation matrix, the Pegboard R+L+B subtest was not included in the selection process.

The Wherry-Doolittle Test Selection method selected the following tests (the standard score regression weights are given in the parentheses): Temperament-I (-.9169); Pegboard-Assembly (.8050); Wonderlic (.6868); Temperament-A (.4558); Temperament-R (-.3162); Temperament-E (.3001); Pegboard-R (.2443); Temperament-D (.1830). Both the Temperament-I and R scores were suppressor variables. The Temperament-I score was highly correlated with both the Temperament-D and E scores; while the Temperament-R score was highly correlated with the Wonderlic score and the Temperament-A

---

[2]H. E. Garrett, Statistics in Psychology and Education (New York: Longmans, Green, 1958) pp. 426-441.

score. The suppressor variable with its negative regression weight removes some of the variance in another predictor which is unrelated to the criterion, This increases the proportion of the variance which is related to the criterion in the other predictor. The shrunken multiple correlation between the battery and the prime criterion was .9213 which was significant at the .01 level of significance. The standard error of estimate for the criterion was .0420.

Concurrent Validity: Secondary Criterion

A biserial correlation and its standard error were computed between each test and the secondary criterion. Each of these correlations was tested for significance. The results are in Table IV. None of them were significant at or beyond the .10 level; only two were significant between the .20 and .10 level of significance; these were the Pegboard-R and B; which correlated with the secondary criterion to the extent of -.3891 and .4430, respectively. These were retained for the cross validation and predictive validity studies.

III. RESULTS IN THE PREDICTIVE GROUP

The second stage of the study involved applicants who were hired as carders. These women were tested at the time of employment. No screening was done however on the basis of

TABLE IV

BISERIAL CORRELATIONS WITH SECONDARY CRITERION IN CONCURRENT GROUP

Variables	r_{bis}	$\sigma_{r_{bis}}$
Wonderlic	.0313	.2738
Pegboard-R	-.3891	.2407
Pegboard-L	.2240	.2628
Pegboard-B	.4430	.2394
Pegboard-R+L+B	.1370	.2700
Pegboard-Assembly	.2811	.2565
Temperament-A	.0228	.2736
Temperament-V	-.0626	.2729
Temperament-I	-.0933	.2719
Temperament-D	.3012	.2540
Temperament-E	.1630	.2679
Temperament-S	.1427	.2693
Temperament-R	.1385	.2695

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

TABLE V

RESULTS IN THE PREDICTIVE GROUP

Variables	Median	Mean	Standard Deviation
Prime criterion	0.174	0.172	0.113
Wonderlic	9.000	9.833	5.490
Pegboard-R	56.500	56.167	5.075
Pegboard-Assembly	137.500	128.750	21.536
Temperament-A	10.000	10.000	2.415
Temperament-I	10.500	10.583	2.218
Temperament-D	9.000	8.583	4.025
Temperament-E	11.000	10.333	3.520
Temperament-R	7.000	7.917	2.499

the test scores. Due to the mobility within the company, many of the carders who were originally tested were transferred to other departments or terminated before they had been on the job fourteen pay periods thus reducing the number of subjects in this stage of the study. Only twelve girls remained in the carding department at the end of the fourteen pay periods.

Comparison of the Concurrent and Predictive Groups

The newly employed carders were given the Pegboard, the Wonderlic, and the Temperament Tests. At the end of fourteen pay periods, the test scores of the twelve remaining carders were correlated with the criterion. The results are contained in Tables V and VI.

Before further analysis and interpretation were attempted, it seemed desirable to discover if the two groups could have come from the same population and were, therefore, comparable. Using the .01 level of significance, the variances and means of the two groups were tested for significant differences as shown in Tables VII and VIII, respectively. Only one of the differences was significant at the .05 level, none at the .01 level. Only one pair of the correlations in Tables III and VI differed enough between the two groups to be significant at the .01 level. This was the relation between the Wonderlic and the Pegboard-R. The correlations of

the Temperament-I with the A and D scores of the same test were the only correlations which differed at the .05 level. Thus the two groups were essentially comparable despite the difference in the nature of the data collection process.

Predictive Validity: Prime Criterion

The tests were then combined into an efficient test battery by use of the Wherry-Doolittle Test Selection method. The test selection method selected the following tests (the standard score regression weights are given in parantheses): the Pegboard-Assembly (.7229) and R (.6929). The shrunken multiple correlation between the battery and the criterion was .7793 which was significant beyond the .01 level of significance. In this case, the standard error was .0714.

Predictive Validity: Secondary Criterion

The biserial correlation between the secondary criterion and the Pegboard-R was .5810 with a standard error of .2626. This was significant at the .05 level. The biserial correlation between the Pegboard-B and the secondary criterion was .6005 with a standard error of .2576. This correlation was significant at the .01 level.

Since the concurrent and predictive groups proved to be similiar a statement concerning the probability of the combined results under the null hypotheses could be made in

TABLE VI

INTERCORRELATIONS OF TESTS AND PRIME CRITERION: PREDICTIVE GROUP

Variables	2	6	7	9	10	11	13	14
1	.90**	.15	-.03	-.03	.07	.10	-.20	-.04
2		-.32	.12	.11	-.28	.09	.00	.41
6			.38	.13	-.44	-.18	-.49	.52
7				.11	-.05	-.22	.01	.34
9					.34	.19	.34	.43
10						.52	.45	-.01
11							-.12	.06
13								-.49

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

KEY: Test	Variable
Wonderlic	1
Pegboard-R	2
Pegboard-Assembly	6
Temperament-A	7
Temperament-I	9
Temperament-D	10
Temperament-E	11
Temperament-R	13
Prime criterion	14

TABLE VII

COMPARISON BETWEEN THE VARIANCES OF THE CONCURRENT GROUP AND THE PREDICTIVE GROUP

Variables	Variance Concurrent Group	Variance Predictive Group	F
Prime Criterion	.012	.012	1.08
Wonderlic	29.127	30.137	1.04
Pegboard-R	50.967	25.743	1.96
Pegboard-B	33.099	9.803	3.38*
Pegboard-Assembly	203.380	447.670	2.20
Temperament-A	7.370	5.833	1.29
Temperament-I	11.417	4.918	2.32
Temperament-D	15.920	16.241	1.11
Temperament-E	11.957	12.389	1.04
Temperament-R	10.903	6.243	1.75

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

TABLE VIII

COMPARISON BETWEEN THE MEANS OF THE CONCURRANT GROUP AND THE PREDICTIVE GROUP

Variables	Means Concurrent Group	Means Predictive Group	Difference
Prime Criterion	.085	.177	.092
Wonderlic	12.333	9.833	2.500
Pegboard-R	54.714	56.167	1.453
Pegboard-B	42.310	44.833	2.523
Pegboard-Assembly	120.955	129.580	8.625
Temperament-A	10.452	10.000	.452
Temperament-I	10.738	10.583	.155
Temperament-D	6.119	8.583	2.464
Temperament-E	8.071	10.333	2.262
Temperament-R	7.881	7.917	.036

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

terms of a compound probability.[3] In the case of the correlation between the Pegboard-B and the secondary criterion, the compound probability was found to be less than .01 while in the case of the correlation between the secondary criterion and the Pegboard-R, it was more than .01.

Because of the low compound probability in the case of the Pegboard-B, an attempt was made to determine a cutting score for this subtest. This was done using the procedure outlined by Kendall and Stone.[4] The best cutting score in the concurrent validity study was 43. The same procedure was applied to the predictive validity group and the cutting score was found to be 43. Therefore, it was recommended that all applicants falling below 43 be rejected.

IV. CROSS VALIDATION STUDY

A second method of estimating the predictive power of a test battery is by cross validation. The regression coefficients from the regression equation of the concurrent validity study were used to compute a composite correlation for the battery and the prime criterion in the second group

---

[3]Baker, op. cit.

[4]W. E. Kendall and C. Harold Stone, Effective Personnel Selection Procedures, (Englewood Cliffs: Prentice-Hall, Inc., 1956).

tested. The composite correlation between the prime criterion and the test battery was .2402. As the predictive validity study found a significant multiple correlation between the prime criterion and the Pegboard-R and Assembly, a composite correlation using the beta coefficients from the predictive study and the concurrent group observations was computed. The composite correlation was .3286.

Caution was required in interpreting all of the correlations between the Pegboard-R and the criteria due to the possible difference between the scores of normal and deaf subjects on the Pegboard-R. On the other hand, the absence of such a difference in the Pegboard-Assembly made it clear that where this subtest correlated with the criteria it made a contribution that could not be attributed to the hearing variable.

CHAPTER V

FINDINGS AND CONCLUSIONS

In order to reduce its cost of doing business, the XYZ company decided to try to reduce its cost of production due to training of new workers by selecting those applicants who would either increase their production more quickly to the production equivalent to the minimum wage or increase their production to above the former plant average. The carding department was picked as the focal point of the study as most employees worked there at one time or another. An examination of the production records was made, and the criteria were picked. The prime criterion was the mean production in dollars actually earned from the tenth to the fourteenth pay period. The secondary criterion was the mean production in dollars actually earned for the first through the fifth pay period.

I. FINDINGS

Three tests were used in the concurrent validity study; they were the Wonderlic Personnel Test, the Purdue Pegboard, and the Thurstone Temperament Schedule. The Pearson product-moment correlation was used to determine the relation between the test scores and the prime criterion. Then the Wherry-Doolittle Test Selection method was used to

select a composite of scores to predict the prime criterion. The result was a composite containing: The Purdue Pegboard Right hand and Assembly scores; the Thurstone Temperament Schedule Impulsive, Stable, Active, Reflective and Dominant scores and the Wonderlic Personnel Test score. This composite had a shrunken multiple correlation with the prime criterion of .92 which was significant at the .01 level.

In a predictive validity study, only those tests which were selected to predict the prime criterion in the concurrent validity study were included. The Wherry-Doolittle Test Selection method was applied again to the data from this group. The Purdue Pegboard Assembly and Right hand scores were found to be the best test battery. The shrunken multiple correlation between two variables and the prime criterion was .78 which was significant at the .01 level. A double cross validation yielded a composite correlation of .24 when concurrent weights were applied in the predictive group and a composite correlation of .33 when predictive weights were applied in the concurrent group.

The biserial correlations between the secondary criterion and the tests proved not to be significant at the .05 level in the concurrent validity study. However, two of the scores correlated enough with the secondary criterion to suggest continuing their use in the validation study. The

predictive study yielded biserial correlations of .5810 and .6005 for these two scores. These were both significant at the .05 level. Upon computation of the compound probabilities, it was found that the null hypothesis for the correlation between the secondary criterion and the Purdue Pegboard Both hands score could be rejected at the .01 level. However, the corresponding hypothesis had to be accepted in the case of the Purdue Pegboard Right hand score. Finally the most efficient cutting score for the Purdue Pegboard Both hands score was determined.

II. CONCLUSIONS

Despite the small size of the two groups, significant results were obtained using a battery consisting of the Purdue Pegboard Right hand and Assembly scores. It would appear that some degree of success in predicting the prime criterion may be possible using these scores. It would also appear that the Purdue Pegboard Both hands score has value as a predictor of the secondary criterion. More research with larger samples must be carried on in order to determine accurately the gains to be expected.

III. SUGGESTIONS FOR FURTHER STUDY

The addition of a job replica would probably increase

the predictive power of the test battery. Such a replica might be constructed using one or two different types of barrettes carded in the plant. The score might be the total number of barrettes carded within a certain time.

Further investigations using larger samples should be undertaken in the area of temperament tests in general and the Thurstone Temperament Schedule in particular. The Active, Vigorous, Dominant, and Social scores of the Thurstone Temperament Schedule show some signs of being of predictive value in this situation. Further study should be made with intelligence tests to see if a curvilinear relation exists between intelligence and production criteria.

BIBLIOGRAPHY

BIBLIOGRAPHY

A. BOOKS

Anastasi, Anne. Psychological Testing. New York: McMillian Company, 1956.

Garrett, H.E. Statistics in Psychology and Education. New York: Longmans, Green, 1958.

Ghieselli, Edwin, and Clarence Brown. Personnel and Industrial Psychology. Englewood Cliffs: Prentice-Hall Incorporated, 1955.

Guilford, J.P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill Book Company, 1956.

Kendall, W. E. and C. Harold Stone. Effective Personnel Selection Procedures. Englewood Cliffs: Prentice-Hall, Incorporated, 1956.

Nemmers, Erwin E. Dictionary of Economics. Patterson: Littlefield, Adams and Company, 1959.

Schnieder, Eugene V. Industrial Sociology. New York: McGraw-Hill Book Company, 1957.

Siegel, Sidney. Nonparametric Statistics for the Behaviorial Sciences. New York: McGraw-Hill Book Company, 1956.

Tiffin, Joseph. Industrial Psychology. Third Edition. Englewood Cliffs: Prentice-Hall, Incorporated, 1956.

Van Steenberg, N. J. "the Thurstone Temperament Schedule," Fifth Mental Measurements Year Book. O. K. Buros, editor. Highland Park: Gryphon Press, 1959.

B. PERIODICALS

Baker, Paul C. "Combining Tests of Significance in Cross Validation," Educational and Psychological Measurement. 12 (1952).

Cuomo, Sylvia, and Herbert W. Meyer. "Validation Information Exchange, #7-077: D.O.T. Code 6-78.632 Floor Assembler," Personnel Psychology VIII (Summer, 1955).

Surgent, L. V. "The Use of Aptitude Tests in the Selection of Radio Tube Mounters," Psychological Monograph, LXI #2 (1947).

U. S. Employment Service. "Validity Exchange, #7-093 D.O.T. Code 7-00.070 Light Bulb Assemblers, "Personnel Psychology VII (Winter, 1954).

C. TEST MANUALS

Thurstone, L.L. Examiner Manual for the Thurstone Temperament Schedule. Chicago: Science Research Associates, Incorporated, 1950.

Tiffin, Joseph. Examiner Manual for the Purdue Pegboard. Chicago: Science Research Associates, Incorporated, 1950.

Wonderlic, E.F. Wonderlic Personnel Test Manual. Northfield: E.F. Wonderlic, 1961.

D. PAMPHLETS AND UNPUBLISHED MATERIALS

The American Economic Foundation. How and Why to Prepare the Functional Operating Report. New York: The American Economic Foundation, 1958.

LeFavre, Robert. "Capitalism in the United States." Read at a meeting of the members of the Midwest Employers Council, Omaha, Nebraska, August, 1961.

APPENDIX

APPENDIX

Regression Equation for Concurrent Validity Study: Gross Score

$$X_c = .0005\ 1/6 + .0154X_1 - .314\ 1/8 + .0097X_{11} + .0197X_Z + .0056\ 1/2 - .0119X_{12} + .0048X_{10} - .383$$

X_c = Estimate of production in terms of deviant wages from the minimum wage.

X_6 = Raw score of Pegboard-Assembly.

X_1 = Raw score of Wonderlic

X_8 = Raw score of the Temperament-I

X_u = Raw score of the Temperament-E

X_7 = Raw score of the Temperament-V

X_2 = Raw score Pegboard-R

X_{12}= Raw score of the Temperament-R

Regression Equation for Predictive Validity Study: Gross Score

$$X_c = .004\ X_6 + .014X_2 - 1.129$$

X_c = Estimate of production in terms of deviant wages from the minimum wage.

X_6 = Raw score of Pegboard-Assembly

X_2 = Raw score of Pegboard-R