

Eastern Illinois University

The Keep

Plan B Papers

Student Theses & Publications

1-1-1967

An Analysis of Industrial Testing for Selection and Placement

Edward D. Dowling

Follow this and additional works at: https://thekeep.eiu.edu/plan_b

Recommended Citation

Dowling, Edward D., "An Analysis of Industrial Testing for Selection and Placement" (1967). *Plan B Papers*. 610.

https://thekeep.eiu.edu/plan_b/610

This Dissertation/Thesis is brought to you for free and open access by the Student Theses & Publications at The Keep. It has been accepted for inclusion in Plan B Papers by an authorized administrator of The Keep. For more information, please contact tabruns@eiu.edu.

AN ANALYSIS OF INDUSTRIAL TESTING

FOR SELECTION AND PLACEMENT
(TITLE)

BY

Edward D. Dowling

PLAN B PAPER

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE MASTER OF SCIENCE IN EDUCATION
AND PREPARED IN COURSE

Education 592

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY,
CHARLESTON, ILLINOIS

1967
YEAR

I HEREBY RECOMMEND THIS PLAN B PAPER BE ACCEPTED AS
FULFILLING THIS PART OF THE DEGREE, M.S. IN ED.

July 26, 1967
DATE
July 26, 1967
DATE

[REDACTED]

ADVISER

[REDACTED]

DEPARTMENT HEAD

ACKNOWLEDGMENTS

I wish to express appreciation to the following persons for their co-operation in the preparation of this study.

1. To Dr. Donald Moler, Department Head of Guidance and Counseling at Eastern Illinois University, for his assistance and recommendations in writing this paper.
2. To Dr. Carl Green, Education Department at Eastern Illinois University, for his recommendations and assistance in finding pertinent reference material.
3. To Dr. William Crane, Education Department at Eastern Illinois University, for his recommendations and information on statistical data analysis.
4. To the Personnel Managers of companies A,B,C,and D in allowing the examination of their test facilities and instruments.
5. To Mrs. Maxine Miller, secretary at Lakeland Junior College, for her help in typing of the paper.
6. To my wife, Helen, who has given support and sacrificed many times throughout my study in Guidance and Counseling.
7. To my sons, Gary and Andy, who have, at a very early age, shown consideration of their father.

Edward Dowling

LIST OF TABLES

Table	Page
I. Testing for Counseling Vs. Testing for Selection	6
II. Tests given by respective companies	20
III. Interview questions and responses	30

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
LIST OF TABLESiii
Chapter	
I. INTRODUCTION	1
Definitions	3
Limitations	5
Related Research	6
II. PRODEDURE IN THE STUDY	15
Guidelines	15
Questionnaire	16
III. ANALYSIS OF TEST MANUAL AND QUESTIONNAIRE	17
General analysis	17
Specific Analysis	21
Specific analysis of Interview Questions	32
IV. SUMMARY AND CONCLUSIONS	36
Problem	36
Limitations.	36
Test Manual Conclusions	37
Interview.	39
Summary and Opinion.	42
APPENDIX.	43
BIBLIOGRAPHY	45

CHAPTER I

INTRODUCTION

Following World War II, industry has increased its use of testing devices for selection and placement. Wartime testing results increased the quality of such instruments due to the availability of large scale testing. "In the Army Air Force, half a million men were tested in four years."¹ It is common knowledge, among those involved in testing, that an increase in the number of testees tends to increase the accuracy of the device. This success indicated a more rapid method of placing and selecting for an industrial organization. With a higher degree of prediction for employee success available, the personnel manager is gradually increasing his choice over guessing. The study made in this paper was made at the level of, and with the cooperation of local personnel managers.

Why is industry concerned about obtaining personnel? According to Rothney, Danielson and Heimann, the following are reasons for the use of testing instruments. "Where the training of employees is expensive - \$5,000 to \$10,000 per man in some cases - the desire of management to reduce failure of employees on the job is understandable. Increasing the number of successes over the number of failures shows up on the profit

¹R. Thorndike, Personnel Selection, (New York, London, John Wiley and Sons Inc., 1949) p. 6

and loss statement, and even a slight improvement over chance selection in industry may justify the use of tests."² In the same book is a list of questions asked by employers of the testing process. A few of these questions follow:

- "1. Would this applicant do better in one phase of our work (mechanical) than another (clerical)?
2. Can you provide us with information that will assist us in the placement of an individual within our organization?
3. How does he get along with others?
4. We are interested in potential leadership - how will he respond to training?"³

The intent here is to place the answers to these questions in the proper frame of reference. Since accurate prediction can never be one hundred percent, it is necessary to be cognizant of what degree of accuracy is possible. The limitations and assets must be kept in mind when making predictions of job success.

With this advent of large scale testing in industry comes a need for accurate measuring devices. The necessity results from more rapid and better employee selection within companies.

Since personnel managers have the responsibility of the testing program, it becomes their duty to acquire knowledge in the selection and use of tests. Though the instruments may be chosen by higher

²J. Rothney, R. Danielson, R. Heimann, Measurement for Guidance, John Guy Fowlkes, (New York, Evanston, and London, Harper, and Row Publishers, 1949) pp. 18-19

³Ibid., p. 8

echelons of management, some technical knowledge remains with the officer in charge of local testing. Tests with poor accuracy of selection will not allow the personnel manager to be reasonably sure the device is performing as intended. One method of discovering inadequacies in a test is by analysis of the test manual which accompanies a test.

It is hoped that the information revealed in this paper will enhance the proficiency of the cooperating industries in making critical analysis of their tests and programs.

One must bear in mind that particular tests may leave a great deal to be desired. The fact remains that they may be the best currently available despite their fallacious nature. Problems may be pointed out in the material but recommendations for replacing tests will not be made to the industries. Reference material will be supplied to those participating in this study as a source of future reference.

Definitions

The definitions provided herein are to establish the writer's terminology when referring to a particular word. In some cases a word may have two meanings. To avoid misinterpretation they must be clarified to establish concrete thought.

1. A-B-C-D: These letters are used to replace company names. The cooperating industries requested anonymity. As a result they were reduced to the above mentioned letters.

2. Validity: Unless otherwise specified the validity will be of predictive nature. This will be the ability of tests to measure what they were intended to measure.
3. Reliability: Items selected in a test are only part of all the items possible to ask in measuring ability. If an individual were retested, his performance should be consistent. The lack of this consistency will not allow one to be certain he has a true sample of this ability.
4. Norms: The group tested to establish test data information. Local norms refer to data compiled by personnel managers of industries using the tests. The tests should be as similar as possible to the individuals used to establish data. Similarities in age, sex, educational level and socio-economic level should be duplicated as nearly as possible.
5. Correlations: "The tests that have high correlations are presumably measuring the same thing to a considerable extent, or are 'saturated' with the same factor to a considerable extent."⁴ A perfect correlation of 1.0 is desired but, theoretically speaking, this is improbable. This 1.0 is a coefficient of correlation, or numerical indication showing the extent of correlation. "Most research workers who have used test scores to predict success in training (or on the job) report coefficients of correlation between test performance and future grades or other criteria of success in the range of $r = .40$ to $.50$. Relationships of such magnitude allow the counselor to improve his predictions over chance only about 12 to 15 percent."⁵ To increase the coefficient is the goal so that one may be more confident in the use of test results.
6. Instrument: Test
7. Device: Test
8. Counselees: Those individuals to whom the tests are given.
9. Counselors: Those individuals who administer and interpret testing devices.

⁴J. Mursell, Psychological Testing, (New York, London, Toronto, Longmans, Green and Company, 1947) p. 384

⁵Rothney, Measurement, p. 229

Limitations

The primary limitation of this paper is the size of the sampling used. Only four local industries are represented. The industrial complex, as a whole, may be either better or worse than the study indicates. These facts must be considered in relation to the title of the paper, which may suggest an extensive study of several corporations throughout the country.

In order to attempt gaining a nationwide representation large companies were selected. Three of the four have regional plants scattered from coast to coast, but the highest concentration is within the eastern seaboard and the Mississippi River area. Only one company is restricted to the State of Illinois.

Since national management is instrumental in test selection the tests are used nationally throughout the various plants within a particular organization. Home offices select these tests, in some cases, and distribute them to local areas unless local job demands require specific skills. In these latter cases area personnel may be required to make a test choice.

The interview questions were prepared by the writer and are only a fragment of numerous guidelines established for complete analysis. This paper can only begin to indicate the problems being questioned and fail to analyze difficulties which are not covered by the questionnaire.

Another factor escapes the study over which there is no control. Some of the manuals are antiquated in relationship to the test. It is possible that new manuals have been published but were not in the

possession of the test administrators at the time of the study. The test may have become more proficient but the lack of up-to-date manuals tend to limit the analytical criticism. Manual corrections may have been made without the writer's knowledge in these instances.

Related Research

Many aspects of testing are involved in the use of testing devices. Testing at its inception was intended to classify the feeble-minded. From this beginning one should consider how testing is adapted to industry and varies from testing done in an educational institution. Where industrial tests are used basically for the benefit of the employer, education attempts to provide information for the student, parent, and institution.

The following information is taken from the test by Rothney, Danielson and Heimann, showing the difference in educational and industrial testing.

"Table I"⁶

Testing for Counseling

Concern for all members of a particular situation, i.e., all students in a given high school regardless of range of performance and characteristics.

Testing for Selection

Concern with limited numbers of applicants for work within a specific organization, with some screening perhaps, involved in the nature of job announcements and specifications.

⁶Ibid., p. 19

Unique concern with one individual at a time. Counseling is an individual affair. Averages or percentages or success is of little comfort to those who are not successful.

Same obligation to all students. Individually, they are very present and a working part of school organizations. Counselors cannot turn them away.

Students are going into a future the dimensions of which are not known. Counselor working with many variables and with many unpredictables.

Testing that improves over chance selection pays off in terms of production, the company's prime concern. Where there is a gain of even one successful employee over failures, testing may more than pay for itself.

In a sense no obligation to any applicant and especially none to those not selected. Future or next steps of those rejected in a hiring situation require no further contact, or the formulation of alternate plans.

Selection made into a defined situation-usually a specific job with dimensions established.

"In any program of personnel selection for a certain number of job specialties, the first step logically and to a certain extent chronologically, is an analysis of the jobs in question to determine the activities which are carried out in those jobs and the circumstances under which they are carried out."⁷

This suggests the necessity of those selecting the tests to be somewhat informed as to the qualifications of the position to be filled. Where this is not possible, those who supervise the position may be consulted when making the choice of an instrument.

"Tests of specific aptitudes are designed for specific populations. For example, one test of ability in art is designed for grades seven and above; one test of mechanical aptitude is to be used for ages eight to twenty-one; a law aptitude test is standardized, of course, for college students and other candidates for admission to law school, regardless of age.

⁷Thorndike, Personnel Selection, p. 12

In any event, whatever the traits or functions to be measured, and whatever the range of ages or school grades for which a particular test is designed, it must be standardized upon a representative group of the population for which it is intended and with which it is to be used. That is to say, each test must be constructed by means of actually sampling the performance of a large group which has been selected in such a way as to insure its being typical of the population for which it is a part."⁸

In this study, three of the four companies A-B-D applied mechanical aptitude tests in personnel selection. The following quotation gives an analytical evaluation of the instruments in general.

"On the whole these tests are statistically reliable. The validity, however, may be questioned; for in evaluating them, psychologists repeatedly comment on the inadequacy of validating criteria. If we regard marks in high school, shop courses, scores of occupational and educational groups (Mechanic versus non-mechanic), and low correlations with tests of general intelligence as criteria, then we can say that some of the available tests in this field have a fair degree of validity for purposes of educational guidance. On the whole, by comparison with tests of intelligence, available tests of mechanical aptitude are inferior in respect to definition of functions to be measured, level of standardization, and predictive value in actual performance."⁹

Companies B, C and D employed the use of clerical tests and the following is an evaluation of these devices.

"While reliability coefficients are generally within satisfactory range, tests of clerical aptitude do not provide sufficient evidence of their general value for the prediction of competence and quality of performance on the job itself. Validity correlations generally fall between .20 and .45. However, as so often happens even when validity correlations are low, the tests are useful in identifying those persons at the higher levels."¹⁰

⁸Frank S. Freeman, Theory and Practices of Psychological Testing, (New York, Henry Holt and Company, 1950) p. 2

⁹Frank S. Freeman, Theory and Practices of Psychological Testing, (New York, Henry Holt and Company, 1955) p. 324

¹⁰Ibid., p. 328

Only companies B and D provided the writer with a manual for the Kuder Preference Test. An indication of the value given to these tests is summed up by Rothney, Danielson and Heimann.

"In this chapter it has been suggested that there are are no valid short cuts to the appraisal of personality, attitudes, interest and behavior of counselees. Examination of the form and content of self-descriptive inventories, records, blanks and projective devices has revealed so many shortcomings that their use by counselors cannot be recommended. (No judgment has been made about their use by clinical psychologists or psychiatrists.) Claims for the value of such instruments are either unsupported or the evidence that is offered is inadequate. It has been suggested that the study of various aspects of the complex behavior of individuals must undergo a long period of descriptive study before valid measurement in this area can be established. Until this is done it has been suggested that the counselor should employ more direct and personal methods of studying the behavior of his counselees by the use of observation, behavior description, interview, and analysis of performance techniques."¹¹

Due to the attention given each test individually, no expansion will be given intelligence testing as a whole. The prime intelligence test, given by all companies in this survey, is thoroughly covered by the next excerpt. This is a complete breakdown of the Wonderlic Personnel Test.

"The easy availability of the tests fail to avoid the possibility of fee employment agencies from purchasing tests. It is probably a rare applicant who is not exposed to the Wonderlic test several times during any given job search; the promotional literature on the test claims that over 6,500 organizations are using it for selection and placement procedures.

For a widely used test it is surprising and very disappointing to find such a paucity of information about the

¹¹ Rothney, Measurement, pp. 315-316

test in the manual.

The references are old - 10 dated in the late 30's and 40's only 6 in the 50's, the most recent being 1958.

The impressive number of 50,000 samples cannot make up for the lack of technical information on validity and reliability. No information is given about the norms, or the kinds of samples employed.

No evidence of validity is given in the manual, although the flat claim is made that the test is shown to be a valid instrument. It is probably true that some studies have shown it to be valid but it should not be necessary for the test user to make up his own literature to find evidence of validity.

It is too bad that the up dated manual was prepared without heading the 1955 American Psychological Association technical recommendations for the preparation of manuals.

It is a widely used, convenient measure of intelligence. It is likely that the test is reliable and valid for a variety of jobs. The lack of informative technical data seems to be its greatest drawback."¹²

Of the four companies in the survey, B, C, and D utilized the General Clerical Test. Its evaluation is made by the Mental Measurements Yearbook.

"Four reviewers wrote evaluations of this test for the Third Mental Measurements Yearbook. They were consistently unenthusiastic. They thought the test was too long, that its content overlapped with that of intelligence tests, that the normative data were inadequate, and that some evidence of the predictive value of its scores should have been provided. These criticisms have, for the most part, been met and dealt with in a revision of the test's manual and in making the test available in 'partial' forms.

The most recent edition of the manual is unusually complete. The prospective user is provided with retest reliabilities, normative data from quite a variety of

¹²Oscar Buros, Mental Measurement Yearbook, (Highland Park, New Jersey, The Gryphon Press, 1965) p. 514

sources, and with the results of a number of validity studies. There is, however, one rather conspicuous omission; the authors fail to describe the process by which the test materials were originally assembled and the process by which they were selected for final inclusion.

Reliability data and interest correlations are reported for a group of 195 seniors in a commercial high school. The reliability data are in the form of retest coefficients. They range in value from .59 to .88 for the subtest scores and from .82 to .91 for the three part scores; these data would suggest that the user of the test probably would not want to make decisions on the basis of individual subtest scores.

Enough validity data are provided to enable one to say that this collection of tests has the potentiality for making a useful contribution to a selection battery. These data take the forms of correlation coefficients and tables which show test performances and job performances.

Norms are presented in the form of percentile rank equivalents. They come from a variety of sources - students in business schools, applicants for clerical positions and employed clerical personnel. A variety of business organizations is represented.

As it stands today with its revised manual, the Psychological Corporation General Clerical Test should be able to hold its own among the best of the clerical aptitude tests; it most certainly is worthy of consideration by every employment office in need of an instrument for selecting or placing clerical personnel. It is probably longer than it needs to be and the separately timed tests may be an annoyance in some employment setting. Only the prospective employer can decide whether these characteristics are really drawbacks."¹³

To continue the individual test research, the Test of Mechanical Comprehension was selected. Companies A, B and D make use of this particular instrument.

¹³Oscar Buros, Mental Measurements Yearbook, (Highland Park, New Jersey, The Gryphon Press, 1953) pp.730-732

"Three forms of this test appeared about a decade ago. A fourth form was added in 1949 extending upward the range of difficulty for male subjects.

The test is intended to measure the ability of an individual to understand various kinds of everyday physical and mechanical relationships. Each form consists of sixty items. Each item includes a picture exhibiting one or more objects, physical situations, or mechanical relationships about which a question permitting a categorical answer is asked. The principles underlying these questions include leverage (perhaps the most frequently represented), force and motion, light, heat and sound, etc. Obviously, the examples used are ones arising out of most people's common experience of physical phenomena rather than coming necessarily from technical training. It is presumably for this reason that previous exposure to formal instruction in elementary physics appears to confer only a slight advantage (about 3 or 4 points).

Reliability (internal consistency) is of the order of .8. For purposes of individual assessment, this is not high but commonly considered as acceptable.

Evidence of validity is given mainly by correlation with job, academic, or training course performance. These range from .3 to .6. Correlations with other tests are of the same order. A limited amount of evidence is available suggesting that this test ranks relatively high in correlating with criteria in comparison with other tests used. While its value will depend in specific situations upon the criterion predicted and the degree of correlation with other tests used, experience with the Test of Mechanical Comprehension would indicate generally it is a useful addition to the stock of measuring devices available for vocational guidance and selection."¹⁴

A second Purdue test, commonly found in industry, covers another field of specialization in the area of electricity.

¹⁴Ibid., p. 767

It is evaluated in the Third Mental Measurement Yearbook by Buros. The copyright date of this particular test is 1942, and is intended to be used for grades 9-16 and adults. Companies B and D utilized this device in their screening.

"No validity studies have been made. The validity is said to rest on the thoroughness of the sampling of electricity and electrical operations in standard textbooks and the fact that particular industrial concepts were incorporated in the construction of the test.

The sampling of the fundamentals of electricity is good, but almost the whole test is devoted to the type of material that an alert high school student can learn from a beginning course in physics.

Any user of the test should realize that the test measures little more than the ground-work in electrical theory. This may well be adequate for most industrial purposes, since a person well grounded in theory can usually be expected to learn fairly rapidly on the job to use an electrician's equipment and to perform an electrician's duties."¹⁵

No mention of reliability is made in the review by those evaluating the test.

Though they were very hesitant to use the Kuder Preference Test, personnel managers from companies B and D sometimes put it to use. The Kuder was the only preference test found locally. It was published in 1948 and underwent changes through 1954 for adapting a comparison of vocational and personal scores.

"Do the scales of the Kuder-Personal actually measure what the titles and explanations indicate?

¹⁵Oscar Buros, Mental Measurements Yearbook, (New Brunswick, New Jersey, Rutgers University Press, 1949) p.671

The method of developing the test is described in the manual. A survey of literature on factorial analysis of personality and interest tests was made. From this survey a list of factors was assembled. Then seven scales, two of which were later abandoned, were selected and items judged to be measures of pertinent factors were assembled for each scale. As far, then, as construction of the test is concerned, the validity of this instrument rests largely upon the judgment of the author that these items actually measure the interest patterns as named and explained.

On the profile sheet is the statement about part E: Preference for directing others. High interest indicates you like situations in which you can influence the thoughts and activities of other people. You like to be in a position of authority! For a student to mark these may mean that he likes the symbols of prestige, but to say he 'likes to be in a position of authority' goes entirely too far.

There is real danger with instruments such as the Kuder Personal that the title of a part will be taken as defined, accurately, measured factor in planning and counseling, when it is not.

In summary, because of the weaknesses discussed above, the reviewer is of the opinion that the Kuder Preference Record Personal is of only limited value. Counselors using it should guard carefully against overinterpretation."¹⁶

The highest correlation found in the test was .57 and the lowest was .30. This was concerning validity coefficients. There was no mention of reliability by the test evaluator, nor was there any word concerning normative data.

¹⁶Oscar Buros, Mental Measurement Yearbook, (Highland Park, New Jersey, The Gryphon Press, 1959) p. 764

CHAPTER II

PROCEDURE IN THE STUDY

Personal contact was made with the personnel managers of the four local industries. Since it was their desire to remain anonymous it is not possible to name the industries. By knowing their own tests, each company is capable of identifying their particular evaluation in the study which would help to shed light on further testing from an objective point of view.

Guidelines

The basis of authority for this paper is described in Rothney, Danielson, and Helmann, Measurement for Guidance. A list of technical recommendations are shown to be essentials required of test manuals. These guidelines were established in 1955 by the American Psychological Association. It was recommended that these standards be met within test manuals as a way to determine the proficiency of a particular test.

Since the writer selected only a few of these standards, a limited analysis is presented here. It is advised that complete informative data can be gained by referring to this more thorough list of specifications.

For a breakdown of each test individually, one can find pertinent data completely described in the various publications of the Mental Measurement Yearbook by Oscar Buros.

Only parts of these analyses were used as they pertained directly to the study, as the analysis was based on copy-right dates, validity, reliability and normative data.

Questionnaire

The questionnaire was intended as a supplement to the direct analysis of instruments. The primary function was an intent to reveal shortcomings of the use and interpretations of testing devices. It will be felt this paper is successful if those participating discover and attempt to correct deficiencies they may become cognizant of as a result of this study.

Each question was intended to elicit self criticism by the test administrators. Most of those interviewed were eager to gain informative data which would otherwise be more difficult to obtain and gain some answers to questions of which they were not previously aware. These questions were not given in a questionnaire form but a short personal interview was conducted following the test manual study. The results of the manual study were not given to the personnel managers at the time of the interview.

CHAPTER III

ANALYSIS OF TEST MANUAL AND QUESTIONNAIRE FINDINGS

The analysis was conducted to find whether or not test manuals were up-to-date with their data. This involved the copyright dates. Validation study was used to find if the tests were capable predictors. Reliability was considered to determine if the tests are true samples of ability. Also norms were taken into account to find if the tests were properly applied to the appropriate populace. Finally, the questionnaire was compiled to elicit self analysis of personal inadequacies regarding understanding in the interpretation of tests and test manuals.

General Analysis

The copyright dates of the manuals ranged from 1940 through 1965. A few of these manuals were revised. In some instances the dates were not found on the manuals but were indicated on the tests. In one particular case, three dates were found on manuals for the same test. The earliest dated 1940 and its latest revision 1961.

Bibliographic information to establish data was found reaching back to 1928. This was true even in the most up-to-date manual for the test.

Of these various tests, the predictive validity ranged from correlations of .02 to .87 with some attempting no correlation whatsoever.

Methods of validation consisted of using curriculum grades as a predictive device to ratings of supervisors on the job. Also teacher's ratings of competence were a method of correlation. To supplement this, correlations were made with various well-known tests assumed to be accurate. Even work records were taken into consideration as a method of establishing validation material.

Few of the tests made any recommendation for the establishment of local validity. They seldom warned of the fact local validity might not correlate with validities established elsewhere.

Within some tests, coefficients of reliability were not listed. Of those available the overall range was from .50 to .98. Methods of establishing reliability were numerous, including test-retest, alternate form, and split-half. The groups used to provide reliability extended from 9th grade students to graduate students in engineering. As often as not these groups were not defined.

Norm groups have been mentioned as being from 9th grade to the graduate level. Sometimes methods of sampling were not noted. Few test manuals mentioned the use of establishing local normative data. Age, sex, educational level, and geographic locale information was found to be sparse. The lowest number of testees was 195 and

the highest in excess of 314,000.

Several times during the study it was found there were no manuals to accompany the tests. In only one case was a test locally devised, therefore, no manual was provided.

Tests given by respective companies

The following list of tests indicate which company in the study makes use of each. Only one is locally devised.

Table II

<u>Company</u>	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>
Wonderlic	X	X	X	X
General Clerical Test		X	X	X
Mechanical Comprehension	X	X		X
Purdue test for Machinists and Machine Operators	X	X		X
Purdue test for Electricians		X		X
Kuder Vocational Inventory		X		X
Flannigan Industrial Test				X
Basic Arithmetic (Locally devised)	X			

Specific Analysis

Test: Wonderlic Personnel Test
E.F. Wonderlic
P.O. Box 7
Northfield, Illinois

Copyright: Three dates were found on these manuals 1945-1955
1961. Two companies were using manuals with the latter of these
dates. The latest edition of the manual had been revised and
enlarged. The original study was made in 1940. The bibliographic
material went back to 1920.

Validity: The author states that the test clearly distinguished
between good and poor against work records after five years.
Also a comment was made that correlation with the Otis Self
Administering Test of Mental Ability was .81 to .87. To establish
validity, 30,000 persons were given the test between the ages of
20 and 65. This is the extent of data revealed on validity.

Reliability: The author believes the test is as good as longer
tests. One method of correlation was immediate retest yielding
.82 to .94. A correction was made for comparison with the Otis,
resulting in a .93 correlation. No mention of the method of
correction was made. An odd-even retest provided a .88 to .94.
Probable error on the 1945 test was 1.3 to 1.6 whereas the 1961
test showed a .16.

Norms: These norms were standardized in business situations with
various work backgrounds. A population of 50,000 was used in six
hundred businesses. One table gave age, sex and educational level

on the 1961 edition. Individuals were used from the 8th grade to college graduates. A total of 27,366 males and 25,988 females were tested.

Miscellaneous: The test is, according to the author, designed for testing adults in business and industrial situations. The test can be used on a 12 minute or unlimited time basis. Company C uses a cut off score of 18 for non professional and 30 for engineers. It is designed to cover areas of geometric analysis, logic, clerical, analogies, arithmetic and verbal. The author believes it to be a test of mental ability compiled by a very complicated formula, which is never mentioned.

Test: General Clerical
Psychological Corporation
522 5th Avenue
New York 18, New York

Copyright: This particular manual was dated 1950, revised from the 1947 version. No bibliographic material was furnished.

Validity: The GCT is correlated with several widely used industrial tests. Correlations of .54 to .83 were gained when compared with the Wonderlic. As correlated with the Science Research Associates Primary Mental Ability test, the yield was .77. From here correlations ran from .75 on such tests as the Test of Primary Mental Ability, .07 on the Pressey Senior Classification test, and .36 to .67 with the Minnesota Clerical. A final correlation was made with supervisor ratings and the range was .34 to .48. Implication is that the validity is predictive in nature.

Reliability: The publishers recommend the test-retest method. This was done, according to the publishers, since speed was an important factor of the test. The time lapse between testings was one month. The authors warn against interpretation of coefficients below .75. Below this level in the test were checking, arithmetic, computation, error location, and grammar. Only alphabetizing, arithmetic reasoning, and spelling exceed this level. Parts are not to be considered separately interpretable, but are meant to be added together to yield scores.

Norms: These are produced from 195 seniors in the Yonkers High School of Commerce New York. Six tables are available in the rear of the manual utilizing regional norms. These include persons from industry, utilities, news media and retail areas. These major titles are reduced to specific job descriptions, such as typing, bookkeeping and filing. In excess of a 10,000 population was used as a background of information.

Miscellaneous: It is stated the test was constructed to measure aptitudes. It is employed in the three major areas of clerical, numerical, and alphabetizing. Numerical is divided into arithmetic computation, error location, and arithmetic reasoning. Lastly, the verbal section of spelling, reading, comprehension, vocabulary, and finally grammar. All the ingredients of an intelligence device are contained in the test.

Test: Test of Mechanical Comprehension
522 5th Avenue
New York 18, New York

Copyright: Three copyright dates were found on the test, these being 1940-1951-1954. Only the 1951 manual had a date stamped on the booklet. The others were found on the tests themselves. Bibliographies contained dates from 1942 to 1950.

Validity: Correlations of .30 to .60 were found in conjunction with the Minnesota Paper Form Board and .37 to .64 with supervisor ratings. The author believed the low correlation enhances the test. The author comments on a higher correlation with college board exams but the coefficient is not listed. On the 1940 test a correlation of .02 to .64 was listed for validity.

Reliability: A split-half reliability of .84 was shown in conjunction with a study of 500 9th grade boys. When used with naval personnel, a correlation of .75 was found with gunners mates, .84 with fire control technicians, and .80 with range finder operators.

Norms: Two tables contained in the manual were general population and professional. Persons from crafts, mechanical, electrical, mechanics and college graduate engineers totaled 314,080. On the 1940 test only 10,000 had been used.

Miscellaneous: The original purpose in designing the test was to determine aptitude for engineering and trade school and the

ability of an individual to understand physical and mechanical relationships for practical situations.

Test: Purdue Test for Machinists and Machine Operators
Science Research Associates
259 East Erie Street
Chicago, Illinois

Copyright: No date was found on the manuals but tests contained the date 1949. Bibliographic data, on which the tests were based ranged from 1920 to 1939.

Validity: No coefficient of validity was given. The author stated that data was not available in sufficient quantity for establishing validity. The basis for the establishment was gained by checking 47 machine shop apprentices with 720 hours of machine shop instruction, then correlating information with teachers ratings of competence. These apprentices were employed by a machine shop company. This is the extent of validity information in the test manual.

Reliability: An odd-even sampling was done yielding a correlation of .80 to .96. Two hundred persons were used in the study of reliability. No other information was given in this area.

Norms: The publishers reiterate norms were compiled from 256 students with machine shop practice and 720 hours of academic training. Age, sex, and geographic location was not tabulated nor was the date of sampling.

Miscellaneous: the publisher indicates the test serves as a terminal achievement exam. It is intended to be used for machine operators alone. Testing requires 50 minutes for administration.

Test: Purdue Test for Electricians
Science Research Associates
57 West Grand Avenue
Chicago, Illinois

Copyright: Dates of copyright on the two manuals were 1942
and 1946. Bibliographic dates began in 1928

Validity: No correlation was given; the author stating a sufficient
sample was not available for validation. The face validity was
based on text books listed in the bibliography, according to the
manual.

Reliability: An alternate for method was used on the 1942 test
resulting in a .91 correlation using 434 electricity students.
In addition 131 sales and service applicants were also provided
from an electrical manufacturing company. On the 1946 test a
.80 to .98 correlation was obtained. In this latter study a 3
day interval and two month interval were used.

Norms: These have been mentioned under validity and reliability,
though age, sex and geographic location was not shown.

Miscellaneous: The publisher calls this test a method of obtaining
achievement level in the craft. The time of administration is not
indicated.

Test: Kuder Preference Record
Science Research Associates
57 West Grand Avenue
Chicago 10, Illinois

Copyright: The copyright date of these test manuals was 1946.

Bibliographic dates were as old as 1922.

Validity: Basically the comments on validity begin with statements indicating individuals score high in their particular skills.

They show musicians score high in music, chemists high in scientific and so on. One would not believe these statements to be of a profound nature. As a comparative device, tables are listed to show the scores made by particular professions in the various classifications.

This would be more valuable to the administrator. The lowest correlation was .295 in opposition to the highest of .57. College grades, high school grades, and occupational achievements were used as comparatives.

Reliability: The reliabilities ranged from .80 to .98 on the reliability table, with a median of .91. Six groups were used including 2,667 adult men in occupations, 166 college students, 500 high school boys, 1,429 adult women in occupations, 101 college women, and 500 high school girls in the junior class.

Norms: The norms indicate 1,858 high school boys in sophomore, junior, and senior classes and 2,005 high school girls in the same levels.

Miscellaneous: Tables are quite extensive in the manual as are formula procedures utilized. This test manual was probably the most extensive of all concerning explanation of the test.

Test: Flannigan Industrial Tests
Science Research Associates
259 East Erie Street
Chicago, Illinois 60611

Copyright: This test was the most recent publication of all those reviewed, the date being 1965.

Validity: The test was correlated with fall semester grades of freshman engineers resulting in figures of .20 to .58. Local validation is recommended to insure selection. Eighteen different tests are part of the total battery. A five year follow up gave a .36 to .65 conclusion. Four hundred thirty eight freshman engineers were used in this analysis.

Reliability: Three tables were listed in the manual. Alternate form methods were used to establish reliability coefficients. The author believes intercorrelations indicate each of the eighteen tests is measuring a job element different from all others. It is suggested that all tests added together would result in a more comprehensive test.

Norms: Seven hundred and one university freshmen were used in the norm sample, plus 3,359 twelfth grade students. This test was a result of Flannigan Aptitude Classification Tests used during World War II on 68,000 aviation cadets.

Miscellaneous: Adaptations of the FACT tests were used in building the FIT tests. The latter tests are shorter and validity of the FACT began at .24 and ended at .60. Warnings are made against the

singular interpretation of the tests, believing total score to be the more proficient. The tests were made for reclassification and upgrading of personnel.

Test: Basic Arithmetic
Locally devised

Copyright: Date of establishment unknown.

Validity: No studies have been made.

Reliability: No studies have been made.

Norms: No studies have been made.

Miscellaneous: There is no manual or data on this test other than it is given to all personnel regardless of job application. The cut off score utilized in this test is 75%.

Interview Questions and Responses

The following table is a list of the interview questions and answers prepared to go beyond the study of test manuals, and gain a more complete study of local testing procedures. Company names are listed as A-B-C-D.

Table III

1. To whom are the tests given?
A All applicants B All applicants C All Applicants
D All applicants
2. Who selects the tests?
A Locally selected B Locally and Home Office C Home Office
D Locally selected
3. Are local personnel consulted for selection?
A Yes B Yes C No D Yes
4. Who administers the tests?
A Pers. Mgr. B Pers. Mgr. C Pers. Mgr. D Secretary
5. Who interprets the tests?
A Pers. Mgr. B Pers. Mgr. C Pers. Mgr. D Pers. Mgr.
6. Has there been a follow up on local validity with a compiling of data?
A No B No C No D No
7. Has there been a follow up on local reliability with a compiling of data?
A No B Yes C No D No
8. What training do you have in testing?
A None B None C None D None
9. Are those who select the tests qualified?
A Yes B Yes C Yes D Yes
10. Do you know technically if a test is good or bad?
A No B No C No D Yes
11. Has there been an establishment of local norms?
A No B No C No D No
12. Why Not?
A No time available B No time available C No time available
D No time available

13. When was this last done?
A Never B Never C Never D Never
14. What do you know technically about personality and preference tests?
A Nothing B Nothing C Nothing D Nothing
15. Are the tests given exactly as directed with no deviations from the directions?
A Yes B Yes C Yes D Yes
16. Are the test manuals read and understood?
A Try to read and understand B Yes C No D Yes
17. Are the test manuals read critically or assumed to be correct?
A Assumed correct B Assumed correct C Assumed correct
D Assumed correct

Specific Analysis of Interview Questions

1. To whom are the tests given?

Tests were given to everyone seeking employment at the four industries. Company C would not allow application for those not completing high school. This same company did not test its engineers locally since it was done in the home office with a test battery requiring one half day. Company B gave the Wonderlic test to all but A, C and D administered the test to office personnel only. Aptitude tests were given to those seeking specific job openings. The Basic Arithmetic Test was given to all seeking employment in company A.

2. Who selects the tests?

Companies A and D selected their tests locally and this was done by the personnel managers alone. The remaining B and C companies had their tests chosen by home office personnel directors. Companies A and B sometimes used local supervisors of a particular job skill when selection is made.

3. Are local personnel consulted for selection?

Companies A and B utilized those in charge of particular skills to help in test selection. These were aptitude tests such as the machine and electrical tests. Companies C and D did not participate in this practice.

4. Who Administers the tests?

Companies A and C used personnel managers for test administration while B and D had this done by secretaries.

5. Who interprets the tests?

All interpretation was done by the personnel managers on the basis of test manual instructions. Scoring of the tests is done by personnel secretaries in all cases.

6. Has there been a follow up on local validity with a compiling of data?

No company had compiled data on validity. Only company D utilized the test scores for information concerning this purpose. This information was forwarded to the home office by company D, but it is not locally available for future use. It should be noted, however, that this company does have direct communication to its home office by means of a WATTS line.

7. Has there been a follow up on local reliability with a compiling of data?

Companies A,C, and D did not retest their employees while company B retested on the request of the employee. Company A had difficulty with retest information. Unions within this company feel that the results may be used to disqualify an employee who is attempting to be upgraded.

8. What training do you have in testing?

Personnel managers of companies A and B had no technical training in testing while the personnel manager of company C stated that it was gained through experience on the job. Company D's personnel manager obtained all information through staff psychologists located in the home office.

9. Are those who select the test qualified?

Company A's representative felt educationally no qualifications had been gained. He believed by the process of eliminating bad tests he had become more qualified. Company C's personnel manager was not sure but felt the home office personnel director was qualified. The reason was that the latter had been hired to upgrade company personnel. Company D's representative said yes since it employed graduate industrial psychologists.

10. Do you know technically if a test is good or bad?

The personnel managers of companies A,B,and C said no while company D's said yes. The latter gave the answer yes since only right and wrong answers could be given.

11. Has there been an establishment of local norms?

There was no case in the study where local norm establishment was attempted.

12. Why not?

The personnel representative of company A said time was not available to do such a study. Also he believed that the local plant was too small and that he knew the employees well enough to not require the study. The representatives of B and C said time was not available. The final personnel manager believed this unnecessary since all information was sent to the home office.

13. When were local norms last established?

Local normative data was not established by any of the

companies involved in the survey.

14. What do you know technically about personality and preference tests?

The representatives of companies A and C do not give these tests while B and D did not feel they knew enough to be qualified in their use. All admitted they knew nothing technically of the interpretation of such tests.

15. Are the tests given exactly as directed with no deviations from the directions?

All participants concurred that exact directions were used and no deviations were made. Only company A's representative said, on second thought, that possible deviations might occur.

16. Are the test manuals read and understood?

The representative of company A stated that he tried to read and understand the manuals. Those from companies B and D said they did read and understand the information. The representative of company C said he did not understand the technical information though he did read it.

17. Are the manuals read critically or assumed to be correct?

In all instances the test manuals were not read critically. Also, all those questioned assumed that those capable of constructing a test must be capable of only true representation. Company D's personnel manager said that he did leave analysis to the staff psychologists.

CHAPTER IV

SUMMARY AND CONCLUSIONS

Problem

The original intent of this study was to determine whether or not industrial testing is lacking in adequate tests and knowledge of these tests. The idea arose after the writer noted many criticisms of educational testing devices. The writer assumed that if educational testing lacked informative data, in all probability the problem would exist in industry. The assumption was made on knowledge of the writer becoming aware that many personnel managers do not have technical backgrounds in testing.

Since the realm of this basic concept of testing is nearly infinite, it became necessary to subdivide the study. Even with this breakdown, the main theme is only a superficial study of its content. This is due to the fact that only a few APA guidelines were used in the survey.

The final result ended in a test manual analysis of copyright dates, validity, reliability, and norm examination. The above was supplemented by a short interview to study knowledge of the test administrators regarding the study material. It is possible that more pertinent questions could have been asked in the interview, but one cannot dwell on the non-existent factors.

Limitations

Some problems exist which limit the study. In some cases, test manuals did not accompany the tests. As a result, specific

reviews of these devices could not be conducted. In the case of the Flannigan Industrial Test, the lack of professional review evidence was not available in the latest Mental Measurement Yearbook. Time is the inherent factor of thorough study for this paper. Problems brought to light are only a facade value of underlying difficulties and should be considered only at that depth. This paper then opens avenues of discovery but does not build the streets.

Test Manual Conclusions

Since the first portion of the study grappled with test manuals, there can be no doubt of the necessity of their existence. APA guideline A.1 makes this recommendation. In the research two tests were found to have no manuals. Without them one cannot examine either the assets or liabilities of the device. A complete lack of data cannot support an instruments use without a previous explanation by authoritative personnel.

Bibliographic dates were found as early as 1920. With the wide developments in industrial mechanics over the past 47 years such a basis can hardly be anything but antiquated.

Copyright dates should be updated as per APA recommendation 2.0. Two of the four companies were found to be using old test manuals of the same test which were utilized by the remaining companies. The latter industries held manuals 16 years newer with more informative data available. It is suggested

that each acquire the latest copies of these manuals.

In those tests which were accompanied by manuals, some lacked any and all mention of validity. Though industry can accept a lower coefficient than education, some of these were low enough to be of little value. Coefficients between .40 to .50 improve group prediction over chance from 12 to 15 per cent. Many were found to be at a lesser plateau.

It would be good practice to establish local validity which was not available to any of the cooperating companies, and which was suggested by few of the publishers. This takes much of the guesswork out of predictive ability, even though it can never be 100% accurate at best. As a word of warning, when validity correlation runs high with another test mentioned, one should first know the predictive ability of the test with which comparison is made. If the comparative device is worthless, you are only proving the inadequacy of our own test.

Not one test indicated what type of validity was referred to as per APA recommendation C.1. No dates of validation gathering were mentioned besides the copyright date.

In some cases reliability information was not present in the test manuals as per APA recommendation D1.0. The sample from which reliability was drawn was also absent in many cases, such as age, sex, educational level, and geographic location. This violates APA recommendation D3.2. One case was cited where immediate re-test was the method of obtaining reliability. In this case the

method would tend to increase the coefficient due to a persons memory of his previous experience with the problems. Again coefficients of reliability, like validity, can be accepted at a lower level but many were low enough to be rejected.

Norms ranging from 9th grade children to graduate engineering students were used. The dates and conditions for sampling were not mentioned. The number ranged from 195 to 314,000. The number of 314,000 makes an excellent sample as opposed to the 195. Even with these ideal sample sizes, the lack of informative data in the manuals destroy nearly all their value. A sample of 195 can hardly be adequate to establish normative information of sufficient quantity. Since these tests were used in industry it does not seem feasible to compare prospective employees with 9th grade children unless the educational achievement sought is the same. The writer frowns on the lack of suggestion for local norms by the publishers. Information about geographic location was absent in many of the instruments. One could hardly compare an Indian from a New Mexico reservation with a psychologist. This is an extreme example but it does explain the point quite readily.

It is suggested that those participating in the study consult bibliographic material at the end of this paper, as the material is easily read without a technical background in testing. This would lead to a better understanding of tests and test manuals.

Interview Conclusions

Initially these tests should be given for the purpose of prediction for job success when they are of an intelligence or

aptitude nature. If one is applying for a clerical position, there is no need to use a test of mechanical ability. If one wishes to be a janitor, there is no need for arithmetic computation. In one particular instance a math test was given to janitorial applicants who use no math in their vocation. Tests should be given for a specific purpose in mind, but not for the act of giving the instrument.

Selection should be made by those qualified, not only by management but also persons who directly supervise the potential applicant. After all, who could be more qualified than the one who knows the capacity necessary for a certain skill? An ideal situation would be where all levels collectively make the choice.

When possible those who interpret tests should administer the instruments. If done, observation of the applicant may shed light on a different interpretation of the results. For example, extreme nervousness by the testee may affect his test score and distort his abilities. One point can make a difference of passing or failing. By lack of observation these factors may be lost.

It is suggested that local validity, reliability, and norm data be formulated or made available to establish clear and concise facts on proper comparative information. Without it one cannot be sure the tests are applicable to the group or to the comparison which is made. It also enables those who follow in the same position to be adequately informed. When unions limit reliability aspects, by trying to prevent retesting, there is little to be done but

attempt to explain why the information would be desirable.

The lack of technical training was admittedly expressed by all participators, but only two of the four felt it was necessary in the performance of their duty. At least a basic course in testing would be desirable. If only the bibliographic material herein is referred to, a great deal of technical information could be gained without formal study. This might help one to recognize his limitations in test analysis.

Perhaps time availability to establish local information is scarce for industry. Yet considering the importance of testing, the time would be well spent. Only 10 minutes per day would enable one to compile data in a short time period.

Knowledge of personality and preference tests was sorely lacking, since no one knew how to interpret them or was aware of the fallacies of interpretation. Again it is suggested a little time reading on these tests would be sufficient to gain an overall picture. A little knowledge is better than none at all if the instrument is used.

Summary and Opinion

From this study general opinions have been formed by the writer. Tests used by industry , locally, appear to be good considering what is currently available for use. The prime problem exists in the lack of technical knowledge possessed by the personnel managers and the lack of informative data contained in the test manuals.

Test manual copyright dates and bibliographic dates were antiquated at the time of the study, suggesting the need for the updating of the material. Correlations of sufficient magnitude should be developed so as to upgrade the technical data of the test manuals and prove the adequacy of the tests.

More interest should be shown in the development of local norms, by both test administrators and publishers. This information is necessary to adapt tests to the locale in which they are to be used.

Administrators would do well to spend a little time in developing their technical knowledge of testing, while publishers should refer to APA guidelines when preparing instruments.

APPENDIX

APPENDIX

A.P.A. Technical Recommendations

- A1. When a test is published for operational use, it should be accompanied by a manual which takes cognizance of the detailed recommendations in this report. ESSENTIAL.
- A2. The manual should be up-to-date. It should be revised at appropriate intervals. ESSENTIAL.
- C1. When validity is reported, the manual should indicate clearly what type of validity is referred to. The unqualified term "Validity" should be avoided unless its meaning is clear from the context. ESSENTIAL.
- D1. The test manual should report such evidence of reliability as would permit the reader to judge whether scores are sufficiently dependable for the recommended uses of the test. If any of the necessary evidence has not been collected, the absence of such information should be noted. ESSENTIAL.
- D3.2 The reliability sample should be described in terms of any selective factors related to the variable being measured, usually including age, sex, and educational level. Number of cases of each type should be reported. ESSENTIAL.

BIBLIOGRAPHY

BOOKS

- Buros, Oscar. Mental Measurement Yearbook. New Brunswick,
New Jersey: Rutgers University Press, 1949
- Buros, Oscar. Mental Measurement Yearbook. Highland Park,
New Jersey: The Gryphon Press, 1953
- Buros, Oscar. Mental Measurement Yearbook. Highland Park,
New Jersey: The Gryphon Press, 1959
- Buros, Oscar. Mental Measurement Yearbook. Highland Park,
New Jersey: The Gryphon Press, 1965
- Freeman, Franks. Theory and Practices of Psychological Test-
ing. New York: Henry Holt and Company, 1950
- Freeman, Franks. Theory and Practices of Psychological
Testing. New York: Henry Holt and Company, 1955
- Mursell, James. Psychological Testing. New York and London:
Longmans, Green and Company, 1947
- Rothney, John W., Danielson, Paul J., Heimann, Robert A.
Measurement for Guidance. New York, Evanston, and London:
Harper and Row Publishers, 1949
- Thorndike, Robert L. Personnel Selection. New York, London:
John Wiley and Sons Inc., 1949