AN INVESTIGATION OF THE PREDICTIVE VALIDITY OF TEST SCORES OBTAINED IN THE COUNSELING SITUATION.

A Thesis Presented for the Degree of Master of Arts by

James Roger Berkshire, B.A.

THE OHIO STATE UNIVERSITY 1947

Approved by:

[Signature]
ACKNOWLEDGEMENTS

The writer wishes to express his gratitude to Dr. Harold A. Edgerton for his guidance and assistance in the development of this thesis; to Dr. John E. Horrocks, Dr. Frank M. Fletcher and Dr. John R. Kinzer for their helpful suggestions and criticisms; and to the Staff of the Occupational Opportunities Service, Ohio State University, for their generous assistance and cooperation in the gathering and preparing of the data used in this study.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II  STATEMENT OF THE PROBLEM</td>
<td>7</td>
</tr>
<tr>
<td>III METHODS AND MATERIALS</td>
<td>9</td>
</tr>
<tr>
<td>IV  RESULTS AND DISCUSSION</td>
<td>16</td>
</tr>
<tr>
<td>V   SUMMARY</td>
<td>28</td>
</tr>
<tr>
<td>VI  IMPLICATIONS AND SUGGESTIONS</td>
<td>30</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>34</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

The accuracy with which college success can be predicted from scores made on aptitude, achievement, and interest tests is obviously a function of the degree of relationship between such scores and college grades. The counselor or administrator making such a prediction compares the test scores made by an individual student with those made by members of a previously tested population, the relationship of whose test scores to their college grades has been empirically demonstrated. This relationship is normally expressed by a coefficient of correlation; and the magnitude of such coefficient determines the degree of accuracy with which predictions from comparable data can be made.

Prediction from test results in this manner makes the assumption that the population of which the present counselee or student is a member is not different from the normative group in the relationship between its test scores and academic success; that the correlation coefficients that express the relationship between the tests scores and college grades of the two populations do not differ by more than the range of their probable errors.

Ideally the normative group has been selected in such a manner as to make the data obtained maximally useful and valid--in the case of tests to be used for the prediction of college success the group is as typical of the college
students in general as the experimenter is able to make it. Thus, if no marked changes take place in the characteristics of subsequent groups entering college, the predictive efficiency of the test for these groups may be assumed to be very similar to that demonstrated for the normative group.

The intrusion of the affects of any selection factor or factors upon the composition of the new group makes the assumptions of similar predictive efficiency or of extrapolatable linearity and homoscedasticity untenable and a re-demonstration of relationships necessary.

In recent years there has been an enormous increase in the amount of counseling and guidance of college students; in much of this, prognosis of the individual's possibilities of success is made from the results of various tests. The accuracy of such prognosis is directly related to the question of whether the students being counseled are typical, as a group, of the normative groups with whom they are being compared.

In the usual university counseling center, the appearance of a student to request guidance, either voluntarily, or because he has been referred by some other branch of the university, has at least the appearance of a selective process, in that other students from the same class and college do not so appear.

In an effort to discover whether the students counseled
at the University of Minnesota were typical of the student body as a whole, Schneidler and Berdie (17) compared the group mean scores made on tests of college aptitude (A.C.E.), English achievement (Co-op), and high school scholarship with those of the total class of which they were a sub-group. In a further study of freshmen in the College of Sciences, Literature and Arts they compared the mean scores of counseled and non-counseled students on five categories of the Minnesota Personality Scales, on the Co-operative Natural Science, Social Science and Mathematics Achievement Tests and on various keys of the Strong Vocational Interest Test. On none of these did they find any significant differences between group means. Therefore they concluded that the students counseled were not different from the general student body, in the factors measured by these tests.

Baller (2), however, investigating the same problem at the University of Nebraska, with a somewhat similar set of tests (Cooperative Contemporary Affairs, Nebraska Test of English Usage, and A.C.E.), found that of 201 freshmen counseled, 61% ranked in the upper 41% of their total class on the test scores obtained. He found also that the mean grade in school of the counseled group was significantly higher than that of the 1695 students not counseled. In both this and the Minnesota study the students had requested counseling on a purely "voluntary" basis. Baller
therefore concludes that, since the earlier study found those students who requested counseling, not significantly different from other students, while his study found them superior, selection factors must differ in different colleges—and that generalization concerning the typicalness of students who request counseling is unsafe.

These two studies dealt with students in school during the years 1939-40-41. The post-war introduction of the counseling program of the Veterans Administration, plus the integration of testing and counseling services into the administrative procedures of many universities, have created a whole new set of factors determinant of whether the student is or is not counseled.

The case records of the Occupational Opportunities Service, the counseling center for Ohio State University, show that students have appeared for counseling for the following stated reasons:

1. For determination of which college to enter.
2. For determination of which major to choose.
3. For determination of whether present curriculum is best choice.
4. For determination of whether to enter college.
5. For determination of whether to leave college.
6. For determination of whether to attempt graduate work.
7. For determination of which area of graduate work to choose.
8. For investigation of why level of academic performance is not up to expectations.
9. For assistance (students who have been dismissed) in setting up new vocational goals.

10. For investigation of whether a proposed transfer of colleges is advisable.

11. For determination of a suitable vocational objective and course of training. (Veterans, under Public Law 16)

12. Students petitioning for readmission after expulsion for failure.

13. Students seeking to enter the University with unsatisfactory scholastic records from other colleges.

14. Students seeking to enter the University with less than normal high school background.

While items 1-2-4-6-7 might quite possibly be selective in the direction found by Baller, items 3-5-8-9-10-12-13 could be reasonably expected to indicate a group tendency toward low scholastic performance, regardless of aptitude.

The student's decision to visit the counseling center may range from purely voluntary - through the mild compulsion of suggestions or referrals by various college administrators - to the absolutely required counseling of potential trainees under Veterans Administration regulations.

The proportions of counselees being counseled for any of the above reasons varies with the progress of the school term, with changes in Veterans Administration and University procedures, and with the amount and types of publicity given to the counseling activities of the agency.

With each shift in the proportions of the various types of cases that make up the total case-load there may be a change in the degree of similarity of the group being counseled to the normative populations from which predictions are
made, or to the college class of which they are a part. Such changes might be discernible from examinations of test and grade data such as those made by Baller and by Schneidler and Berdie, or they might not.

Durflinger (8), after factor analysis of the results of a correlation experiment involving tests of college aptitude, achievement, English, the Bernreuter Personality Inventory, the Providence Music Test, and personal data—with college grades as the criterion—concludes that approximately one-half of the variability of college grades is accounted for by current test batteries, that one-quarter is unreliability of the college grades, and that the remaining quarter is specific, unmeasured factors (or factor).

If the students who are received for counseling include in their number many of those whose variability from expected college performance is related to these unmeasured factors, the predictive efficiency of the tests used might be appreciably decreased, even though tests of typicalness through mean score and mean grade data showed no significant differences from the balance of their college class. Certainty as to predictive validities of the tests could only be reached through re-demonstration of their correlation with the criterion.

In view of the wide variety of selection factors that may affect the composition of the population being counseled in any counseling center at any time such re-demonstrations of validity would seem almost a constant responsibility.
CHAPTER II
STATEMENT OF THE PROBLEM

It is the purpose of this study to investigate the actual relationships between the test scores and the scholastic performance of a group of students counseled at the Occupational Opportunities Service of Ohio State University; and to compare these relationships with those ordinarily presumed to exist for the tests used.

The study is confined to students whose problems related to the College of Engineering; since a relatively large number of counseling cases deal with such problems—and since the group of tests thought to be applicable to prediction of success in engineering is broader, and at the same time more specific, than for other colleges.

It is felt that such an investigation will serve as a check upon the validity of those counseling procedures which involve prediction from individual test scores; and that it will thus, indirectly, throw some light on the question raised by Baller and by Schneider and Berdie, as to the typicalness of counseled populations.

It is not the purpose of this study to identify, for use, those tests having the greatest predictive efficiency: the hypothesis previously set forth—as to the frequently changing character of counseled populations—would need to be disproven before tests that may be identified as having predictive superiority in this study could be assumed to

- 7 -
hold that superiority with other populations at other times.
CHAPTER III
METHODS AND MATERIALS

A. The Experimental Group.

The case files of Occupational Opportunities Service were examined, and all cases counseled during the fourteen months from March, 1945 to April 1946, inclusive, that might conceivably have been in, or were considering entrance into the College of Engineering, were extracted.

A total of 213 such cases were found—and their names were entered on 3 x 5 cards, together with the date on which counseling was initiated.

A search of the files of the College of Engineering in October, 1946, revealed that of these; 74 had not entered the College of Engineering; 9 had entered, but had withdrawn before completing one quarter of attendance; 6 who had been in the College previous to the war did not re-enter after counseling; 10 had transferred to other colleges after counseling and had had too little work in Engineering to provide adequate criterion data for the study; 12 were in Engineering College, but had entered too late to provide adequate criterion data.

This left 102 cases that satisfied the requirements of the study, in-so-far as criterion data was concerned.

B. The Criterion.

Ohio State University operates under the quarter system,
in which the calendar year is divided into four approximately equal three-month school terms. Grades in courses are assigned at the end of each quarter. For purposes of determining the students' standing, and their eligibility for continuing in the University, the standard letter grades that are awarded are converted to honor points. This is done by multiplying the number of hours credit received in a course by the numerical value of the grade received; A being 4, B-3, C-2, D-1, and E being 0. The students' cumulative standing may then be determined at any time by dividing the number of academic hours attempted into the total honor points earned.

Criterion scores for this study are the cumulative point-hour ratios of the students involved, taken for the two quarters most contiguous to the time at which they were counseled. Since some of the students were forced to take certain non-credit courses to make up for high school deficiencies; and since these courses were felt to be non-comparable to the regular academic subjects and were not included as part of the criterion data; the total number of credit hours defined in this study as "two quarters" of work varies, for individuals, from 19 to 37.

Since counseling, in individual cases, may take place either before entrance into a college, during the individual's college term, or after leaving school, the grade record most contiguous in time to the time of counseling was used--
with preference, where a choice was possible, given to the two quarters immediately after counseling.

Of the cases ultimately found useful for this study the following relationships between time of counseling and time during which performance was measured obtained:

32—two quarters following counseling.
24—current quarter and one following.
7—current quarter and one previous.
3—two immediately previous quarters.

It requires three quarters for a student carrying a normal load of courses to complete the work of one class level and pass to the next. Thus the quarter's work in which a student is engaged is identified as Freshman 1, 2, or 3; Sophomore 1, 2, or 3; etc.

The criterion scores of this group were obtained from the following periods in their college careers:

<table>
<thead>
<tr>
<th>Class</th>
<th>Quarters</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman</td>
<td>1-2</td>
<td>37</td>
</tr>
<tr>
<td>Freshman</td>
<td>2-3</td>
<td>14</td>
</tr>
<tr>
<td>Fr.-Soph.</td>
<td>3-1</td>
<td>7</td>
</tr>
<tr>
<td>Sophomore</td>
<td>2-3</td>
<td>3</td>
</tr>
<tr>
<td>Soph.-Jr.</td>
<td>3-1</td>
<td>3</td>
</tr>
<tr>
<td>Junior</td>
<td>2-3</td>
<td>1</td>
</tr>
<tr>
<td>Jr.-Snr.</td>
<td>3-1</td>
<td>1</td>
</tr>
</tbody>
</table>

Using the classifications current in the counseling center; 23 of these were compulsory referrals under Public
Law #16, 25 requested counseling under Public Law #346, and the remaining 18 were classed as not official Veterans Administration referrals. These latter may or may not have been veterans—but at the time of counseling they could not satisfy the requirements of the Veterans Administration as to presenting evidence of eligibility under the law—and thus were classified as non-veteran cases.

C. The Tests.

It is the practice, at Occupational Opportunities Service, to leave to the discretion of each counselor the assignment of the tests appropriate to each case. In the prediction of success in engineering such choices would most commonly include a test of college level academic ability, interest tests, a test of mechanical comprehension, a test of spatial relations, and the most appropriate mathematics test. The tests in most frequent use at the center to investigate these factors are the following:

1. Ohio State Psychological Examination. (Form 22)
2. Strong Vocational Interest Blank for Men.
4. Test of Mechanical Comprehension, Form BB. (Bennett and Fry)
5. Revised Minnesota Paper Form Board Test. (Likert and Quasha)

The 102 cases that satisfied requirements as to adequacy of criterion data were examined in order to find the
largest possible group with the greatest number of tests in common.

The final population used in the study consists of sixty-six cases for which scores were available on the Ohio State Psychological Examination, the Strong Vocational Interest Blank, the Test of Mechanical Comprehension, and the Minnesota Paper Form Board. (These will be herein after referred to as the O.S.P.E., the Strong, the Bennett-Fry, and the Paper Form Board)

The scores on the O.S.P.E. were expressed in percentile ranks and the normative group from which these were drawn consists of Ohio State University Freshmen.

The scores on the Strong are the letter rankings devised by Strong, ranging from a high of A, to a low of C, and expressing degrees of similarity to the interests found by Strong to be typical of various professional groups. (20) The scoring method devised by Dunlap (13) is in use at the center; scores were available on the following interest areas:

1. Arts-Science Technologist. (Strong Group I)
2. Physician
3. Engineer (Strong Group II)
4. Chemist
5. Math-Science Teaching (Strong Group IV)
6. Social Service (Strong Group V)
7. Salesmen (Strong Group IX)
8. Business (Strong Group VIII)

9. Writers and Lawyers (Strong Group X)

The scores on the Bennett-Fry and the Paper Form Board were expressed in percentiles that were based on the performance on the tests of Engineering School Freshmen. (4) (14)

D. The Procedures.

To facilitate transferral of the scores to Hollerith cards, the various scores were translated to coded scores in the following manners:

Percentile scores were coded by decile intervals, with all scores falling within each interval being given the same coded value. These ranged from 0 to 9.

The letter scores on the Strong test were translated to coded values from 0 to 5, for letter scores from C through A, respectively.

Point-hour ratios were coded in 8 half-point steps, ranging from zero, for the interval from .00 to .49, to seven for the interval from 3.50 to 4.00.

While it was recognized that correlations obtained from percentile scores might be of somewhat smaller magnitude than they might be if the data were converted to some form more representative of the normal curve (see Bennett) (3), and that the coarsened groupings involved in coding by decile intervals might have a similar effect, the procedure was adhered to because percentile scores are what the counselor
deals with and their deficiencies in representing a normal
distribution are part of the counseling situation.

After coding, the data were punched into Hollerith
cards as follows:

<table>
<thead>
<tr>
<th>Column</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2-3</td>
<td>Case identification number</td>
</tr>
<tr>
<td>4</td>
<td>Type of case (non-vet., P.L. 346, P.L. 16)</td>
</tr>
<tr>
<td>5</td>
<td>Strong -- Arts-Science-Tech.</td>
</tr>
<tr>
<td>6</td>
<td>&quot; Physician</td>
</tr>
<tr>
<td>7</td>
<td>&quot; Engineer</td>
</tr>
<tr>
<td>8</td>
<td>&quot; Chemist</td>
</tr>
<tr>
<td>9</td>
<td>&quot; Math-Science Teaching</td>
</tr>
<tr>
<td>10</td>
<td>&quot; Social Service</td>
</tr>
<tr>
<td>11</td>
<td>&quot; Salesman</td>
</tr>
<tr>
<td>12</td>
<td>&quot; Business</td>
</tr>
<tr>
<td>13</td>
<td>&quot; Writers--Lawyers</td>
</tr>
<tr>
<td>14</td>
<td>O.S.P.E. Total Score</td>
</tr>
<tr>
<td>15</td>
<td>Bennett-Fry</td>
</tr>
<tr>
<td>16</td>
<td>Paper Form Board</td>
</tr>
<tr>
<td>17</td>
<td>Point-hour Ratio</td>
</tr>
<tr>
<td>18</td>
<td>Item count</td>
</tr>
</tbody>
</table>

The correlations between all variables were then ob-tained by the use of the Hollerith machine and the L-Method
as described by Toops. (23)

Following this, the tests that best predict the criterion,
and their multiple correlations with the criterion, were
obtained by the use of the Wherry-Doolittle Test Selection
technique, as described by Stead and Shartle. (19)
CHAPTER IV
RESULTS AND DISCUSSION

The distributions of the scores on each of the variables and the criterion are set forth in Table 1, on page 17. It will be noted that the scores on Engineer, O.S.P.E., Bennett-Fry, and Paper Form Board—which are normally used in predicting success in engineering—have a marked negative skewness.

37 of the cases had scores of A or B-plus on the Engineers scale, with the mean being above B.

40 of the cases had O.S.P.E. scores above the 60th centile, with the mean score at the 58th centile.

43 of the cases had Bennett-Fry scores above the 60th centile, with the mean score at the 62.6th centile.

50 of the cases had Paper Form Board scores above the 60th centile, with the mean score at the 68th centile.

The group being investigated is obviously a selected one in-so-far as the capacities measured by these tests are concerned in relation to the normative groups with which their scores compare them.

On the other hand, the distribution of scores on the criterion is far more normal, with the mean point-hour ratio at 1.91, or approximately C-minus. There is no exact evidence as to the mean point-hour ratio in the College of Engineering, but since a ratio of 1.8 must be maintained by the student to avoid being dropped, it seems reasonable to believe that
it must be above 2.0. If this be true, then the experimental group is either somewhat poorer than, or at least no better than the college group of which they are a sub-group;

**TABLE I**

**DISTRIBUTION OF SCORES**

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Mean</th>
<th>S. D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>5</td>
<td>10</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>27</td>
<td>8</td>
<td>13</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>12</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>16</td>
<td>6</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>6</td>
<td>6</td>
<td>13</td>
<td>16</td>
<td>12</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>11</td>
<td>15</td>
<td>15</td>
<td>12</td>
<td>7</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>6</td>
<td>9</td>
<td>18</td>
<td>16</td>
<td>11</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>6</td>
<td>6</td>
<td>13</td>
<td>11</td>
<td>16</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>8</td>
<td>17</td>
<td>22</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td>5</td>
<td>9</td>
<td>10</td>
<td>13</td>
<td>8</td>
<td>5.80</td>
<td>2.43</td>
</tr>
<tr>
<td>11.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>10</td>
<td>7</td>
<td>11</td>
<td>14</td>
<td>11</td>
<td>6.26</td>
<td>2.30</td>
</tr>
<tr>
<td>12.</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>15</td>
<td>10</td>
<td>19</td>
<td>6.79</td>
<td>2.18</td>
</tr>
<tr>
<td>13.</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>16</td>
<td>21</td>
<td>12</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**KEY**

2. Physician  9. Writers & Lawyers  
3. Engineer  10. O.S.P.E.  
6. Social Service  13. Point Hour Ratio  
7. Salesman

- 17 -
while being superior to the normative populations on the independent variables. Without further evidence this would seem to indicate that the total engineering college population is markedly superior to the normative groups of the tests.

An examination of the correlation data (Table 2, page 19) discloses that of these four variables only the O.S.P.E. scores are significantly correlated with the criterion, having an r of \( .409 \pm .103 \) and being significant at the 1% level of confidence.

Stuit, (21) in a series of ten studies of the fluctuation of correlations between O.S.P.E. scores and college grades of all Teachers College freshmen at the University of Nebraska, during a five year period, found the first semester correlations ranging from .43 to .62, and the second semester correlations ranging from .41 to .58. The numbers of cases in each study varied from 153 to 245.

Amstutz (1) found a correlation of .607 between first quarter point-hour ratio and O.S.P.E. for 462 unselected College of Education freshmen at Ohio State University.

Garrett (9), using the scores of 200 students who had taken the O.S.P.E. in high school between 1935 and 1940, and who were attending 52 different colleges, found a correlation of \( .608 \pm .030 \) with their college grades.

Hartson (10), investigating the relationship between O.S.P.E. and first semester scholarship of 331 unselected Oberlin College freshmen, found a correlation of .640.
<table>
<thead>
<tr>
<th></th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
<th>8.</th>
<th>9.</th>
<th>10.</th>
<th>11.</th>
<th>12.</th>
<th>13.</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Engineer</td>
<td>.119</td>
<td>.089</td>
<td>.100</td>
<td>.760.xx</td>
<td>-.147</td>
<td>.644.xx</td>
<td>-.341.xx</td>
<td>-.144</td>
<td>.170</td>
<td>.120</td>
<td>-.175</td>
<td>.261.xx</td>
<td></td>
</tr>
<tr>
<td>5. Chemist</td>
<td>.121</td>
<td>.066</td>
<td>.074</td>
<td>.064</td>
<td>-.028</td>
<td>-.466.xx</td>
<td>-.589.xx</td>
<td>-.267.xx</td>
<td>.229</td>
<td>.274.xx</td>
<td>.258.xx</td>
<td>.214</td>
<td></td>
</tr>
<tr>
<td>6. Math-Sci-Teach</td>
<td>.123</td>
<td>.112</td>
<td>.107</td>
<td>.121</td>
<td>.124</td>
<td>.480.xx</td>
<td>-.378.xx</td>
<td>-.258.xx</td>
<td>-.738.xx</td>
<td>-.156</td>
<td>-.029</td>
<td>-.077</td>
<td></td>
</tr>
<tr>
<td>7. Social Service</td>
<td>.124</td>
<td>.096</td>
<td>.108</td>
<td>.073</td>
<td>.098</td>
<td>.085</td>
<td>-.005</td>
<td>.152</td>
<td>-.421.xx</td>
<td>.026</td>
<td>.028</td>
<td>-.162</td>
<td></td>
</tr>
<tr>
<td>8. Salesman</td>
<td>.122</td>
<td>.103</td>
<td>.107</td>
<td>.110</td>
<td>.061</td>
<td>.106</td>
<td>.124</td>
<td>.239</td>
<td>.141</td>
<td>-.340.xx</td>
<td>-.266.xx</td>
<td>-.242</td>
<td></td>
</tr>
<tr>
<td>9. Business</td>
<td>.118</td>
<td>.074</td>
<td>.067</td>
<td>.121</td>
<td>.115</td>
<td>.115</td>
<td>.121</td>
<td>.117</td>
<td>-.483.xx</td>
<td>-.114</td>
<td>-.232</td>
<td>-.092</td>
<td></td>
</tr>
<tr>
<td>10. Wrt. and Law.</td>
<td>.109</td>
<td>.080</td>
<td>.074</td>
<td>.120</td>
<td>.117</td>
<td>.066</td>
<td>.108</td>
<td>.121</td>
<td>.098</td>
<td>.196</td>
<td>.084</td>
<td>.138</td>
<td></td>
</tr>
<tr>
<td>11. O.S.P.E.</td>
<td>.103</td>
<td>.122</td>
<td>.119</td>
<td>.122</td>
<td>.135</td>
<td>.121</td>
<td>.124</td>
<td>.110</td>
<td>.124</td>
<td>.119</td>
<td>.269.xx</td>
<td>-.272.xx</td>
<td></td>
</tr>
<tr>
<td>12. Bennett-Frey</td>
<td>.121</td>
<td>.118</td>
<td>.118</td>
<td>.120</td>
<td>.116</td>
<td>.124</td>
<td>.124</td>
<td>.116</td>
<td>.117</td>
<td>.123</td>
<td>.108</td>
<td>.600.xx</td>
<td></td>
</tr>
</tbody>
</table>

xx—significant at 1% level of confidence
x—significant at 5% level of confidence
lower quadrant shows probable errors of the correlations.

- 19 -
It will be noted that the correlation obtained in this study is at the extreme lower end of the range of correlations cited. Since these studies did not specifically deal with grades in Engineering College one might be inclined to assume that this lower correlation is normal for Engineering subjects. Holcomb and Laslett (11) however, report a correlation of .55 between scores on the A.C.E. Psychological Examination and Engineering grades, while Vaughn mentions a correlation of .58 between scores on the Yale Scholastic Aptitude Test and Engineering grades. Since these tests are roughly comparable to the O.S.P.E. it seems more probable that the smaller correlation obtained in this study is a product of the restriction of range that is evident in the distribution of O.S.P.E. scores for this counseled group.

None of the other three tests normally used for prediction in engineering have correlations with the criterion of sufficient magnitude to be considered significant at the 5% level of confidence.

The largest of the three correlations was that of the Strong Engineer scale, which has a correlation of .195 ± .119 with the criterion; .243 being necessary to reach significance at the 5% level.

While Strong (20) does not expect that Strong scores will correlate highly with scholarship, and while most counselors would not admit to using interest test scores to
predict success in any field; nevertheless, in the counseling situation, the counselor is apt to weight the score on the Strong Engineer scale rather heavily into the decision as to whether this individual does or does not "belong" in Engineering. The assumption that "belonging" in a field means better work there than in a field in which the individual does not "belong" is almost unavoidable; and the further assumption that degree of interest (or degree of belongingness) forecasts degree of performance follows quite naturally.

Brush (6) has stated, after a survey of the existing literature,

It appears that correlations in the thirties can be expected between scores on the Engineer scale of the Strong Vocational Interest Blank and Engineering grades.

and Holcomb and Laslett (11) report such a correlation (.322). Berdie (5), however, found a correlation of .13 between such scores and the honor-point ratios of 43 students after one year of work. Segal and Brintle (18) obtained a correlation of .14 with mathematics and science grades on 93 cases.

The correlation of .195 found in this study, could be obtained by chance from a universe where the true value was either .13 or .32 and cannot therefore be assumed to be different from them. Since the distribution of scores on this scale is so negatively skewed in this study, it seems probable that high scores are sufficiently typical of students
in Engineering colleges as to nearly eliminate the value of the scale, in its customary usage, as a predictor of the comparative success of individuals.

While the Bennett-Fry Test of Mechanical Comprehension is widely used for the prediction of success in college engineering courses, and is among the tests recommended as a "good test battery" for such purposes by the Veterans Administration (26), there is not very much research evidence to support such use. The test manual reports validity coefficients from .3 to .6 with "success at engineering type occupations .......... of a military nature." McDaniel and Reynolds (16) report a correlation of .24 with ratings of 147 males in a National Defense Training Program. Correlations with A.C.E. scores of .10 and .19 are reported in the test manual (4), and a correlation of .367 with A.C.E. scores is reported by Traxler. (24) This latter compares closely with the correlation of .359 between Bennett-Fry scores and O.S.P.E. scores found in this study; a correlation of sufficient magnitude, in comparison to the Bennett-Fry's correlation with the criterion, as to raise the question of whether the Bennett-Fry has any predictive relationship to the criterion that is distinct from general scholastic aptitude (or intelligence) as measured by the O.S.P.E.

Investigation of this question by partial correlation technique reveals that, when O.S.P.E. is held constant, the correlation of the Bennett-Fry scores with point-hour ratio
is reduced to .018. We may therefore conclude that, for this group, the test was without independent predictive value.

The Revised Minnesota Paper Form Board Test is likewise recommended by the Veterans Administration (26) and is in very wide usage in counseling centers. Brush (6) found correlations between scores on the test and engineering grades ranging from .175 to .426 for successive freshman classes. Berdie (5) found a correlation of .22 with the first year grades of 156 engineering students. Stuit and Lapp (22) obtained correlations of .258 and .261 with the scores of two groups of students (N's were 80 and 75) on a Cooperative Physics Achievement test. Bryan (7) found correlations with grades ranging from .10 to .33 for 1000 Pratt Art School freshmen, who had been split into four groups according to their major. Traxler (24) found the scores on this test correlated .393 with those on the Bennett-Fry, as against the correlation of .50 between them found in this study.

When partial correlation procedures were applied, it was found that, with O.S.P.E. held constant, the correlation of the Paper Form Board with the criterion is reduced to .007; indicating that, for this group, this test also had no independent predictive value.

Since the correlations of the tests normally used for prediction of success in engineering were found to be some-
what discouraging, the Wherry-Doolittle Test Selection technique was applied to the total data in order to determine whether any other tests had greater validity for prediction of the criterion.

This technique selects, by use of the correlation and intercorrelation data already obtained, those tests which together give the maximum multiple correlation with the criterion. The order in which the variables were selected, the amounts that each added to the multiple correlation, the beta weights for each variable, and the final multiple correlation are reproduced in Table 3, on page 25.

It is interesting to note that the Writers-Lawyers scale on the Strong Interest Test adds the greatest amount to the multiple correlation after the O.S.P.E. This was a result of its correlation of .348 with the criterion, while its correlation with the O.S.P.E. scores was but .196—indicating a relationship with grades that is partially independent of those things measured by O.S.P.E.

A relationship between scores on this Strong scale and scores on a college aptitude test (A.C.E.) was demonstrated by Long. (15) Tussing (25), using item analysis techniques, succeeded in constructing a scoring key for the Strong test, the resultant scores from which correlated .42 with college grades—and reports that the great majority of the items included were also included in the Writers-Lawyers scale.

It appears that the independence of the predictive
values of this scale should be further investigated -- particularly since, from the nature of the items included

TABLE 3

RESULTS OF APPLICATION OF WHERRY-DOOLITTLE Technique

TO THE CORRELATION DATA

<table>
<thead>
<tr>
<th>TEST</th>
<th>( R )</th>
<th>INCREASE</th>
<th>BETA WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. O.S.P.E.</td>
<td>.409</td>
<td>.409</td>
<td>.359</td>
</tr>
<tr>
<td>2. Writers-Lawyers</td>
<td>.480</td>
<td>.071</td>
<td>.671</td>
</tr>
<tr>
<td>3. Math-Sci-Teach</td>
<td>.521</td>
<td>.041</td>
<td>.369</td>
</tr>
<tr>
<td>4. Engineer</td>
<td>.523</td>
<td>.002</td>
<td>.566</td>
</tr>
<tr>
<td>5. Chemist</td>
<td>.561</td>
<td>.038</td>
<td>-.367</td>
</tr>
<tr>
<td>6. Social Service</td>
<td>.587</td>
<td>.026</td>
<td>.307</td>
</tr>
</tbody>
</table>

in it, and the low relationships with other predictors evident in this study, it may well be predicting within the 25% of the total variance not now accounted for by the usual college battery. (See Durflinger, page 6).

The multiple correlation of .587 is of fairly respectable magnitude, compared to the usual battery of college predictors; but the ingredients in our prediction battery are certainly not those that would have been arrived at by processes of logic or by face validity of the tests.

In fact, beyond the first two tests selected, there seems considerable probability that the selection of the other four Strong scales by the Wherry-Doolittle technique may be a function of the internal inter-relations of the various Strong scales. In other words, the question of the
applicability of the Wherry-Doolittle technique to the selection of more than one predictor from multiple-category tests of the type of the Strong interest test (or the Kuder Preference Record) is being raised—since the scores obtained by an individual on the various scales are not completely independent of the scores he obtains on other scales. The response of 'Like', 'Indifferent', or 'Dislike' to an item on the Strong test may be counted on several scales, thus creating inherent positive intercorrelations between some scales and inherent negative intercorrelations between others. Inspection of the correlation matrix (Table 2, page 19) makes this evident, with the intercorrelations of .750 between Engineer and Chemist, and -.739 between Writer-Lawyer and Math-Sci-Teaching being the most outstanding examples.

Since these high intercorrelations are produced by the fact that the same questions (and answers) are common to two tests, rather than that the two tests are, to the extent of their correlation, measuring a common area by the use of differing items—it seems highly probable that the working out of the Wherry-Doolittle would be consistently and grossly affected thereby, and that the resulting battery of predictors is misleading and the calculated multiple correlation with the criterion spurious.

The application of the Wherry-Doolittle technique to the data in this study, beyond its selection of the C.S.P.E. and the Writer-Lawyer tests, thus serves principally to
demonstrate rather conclusively that the Bennett-Fry and Paper Form Board were without independent predictive validity--since the calculations reached the maximum multiple correlation without having included them.
CHAPTER V
SUMMARY

An investigation was made of the relationships between the test scores of 66 male students who were counseled at Occupational Opportunities Service and their grades in Engineering College at Ohio State University. As a result, the following conclusions were reached:

1. The scores of the experimental group were markedly superior to those of the normative groups on the four tests used for prediction in Engineering. (O.S.P.E., Strong Engineer Scale, Bennett-Fry Test of Mechanical Comprehension, Revised Minnesota paper Form Board.)

2. The grades of the experimental group were not superior to those of non-counseled students.

3. The correlation of O.S.P.E. scores with grades for this counseled group (.409) is lower than is normally expected.

4. The correlation of the Strong Engineer scale with grades is too low (.195) to be independently useful for prediction.

5. The correlation of the Bennett-Fry scores with grades, when O.S.P.E. is held constant, (.018) indicates that, for this group, the test had no predictive value independent from general scholastic
aptitude.

7. Scores on the Strong Writers-Lawyers scale had a correlation with grades that was significant at the 1% level of confidence (.348), and, used in combination with O.S.P.E. scores, added .071 to the multiple correlation.

8. Evidence cited from the intercorrelation data and from analysis of the Strong test is taken to indicate that the Wherry-Doolittle Test Selection Method is not applicable to more than one scale of multi-scale tests in which the scales have inherent intercorrelations.
CHAPTER VI

IMPLICATIONS AND SUGGESTIONS

This investigation seems to support Baller's conclusion, i.e. that generalization concerning the typicalness of students who request counseling is unsafe. The relations of test scores to the scores of normative groups and to school grades are different from those found by either Baller or by Schneidler and Berdie. This further emphasizes the necessity for frequent re-investigation of test validities for counseled groups.

Validities that have been demonstrated for other groups, under other circumstances, can only be tentatively accepted for use, and used with the greatest of caution and conservatism.

The selective factors peculiar to this experimental group; the variation in their reasons for requesting counseling, the differences in their class ranks, the variations in time relations of criterion data and test data, all the factors that in the average study would be considered statistically "unsanitary"; are factors that are native to the counseling situation. These are the people with whom the counselor deals—and this is the situation which should strictly limit the wise counselor's confidence in his tests as predictors.

The results found in this study for the Bennett-Fry and Paper Form Board demonstrate the futility of many studies in
the literature. The measurement and re-measurement of the predictive validities of so-called special ability tests, without holding general intelligence or scholastic aptitude constant, is all too common. The correlations thus obtained are then looked on as an indication of the relation of that "special ability" to success in the field; and scores on the test are considered predictive in a manner distinct from the predictions from scholastic aptitude scores. A high score on the "special ability" tests, in addition to a high score on the general aptitude tests, is assumed to make success doubly certain. If, as was true in this study, the "special ability" tests are merely measuring one part of general ability, then the above assumption can only lead to error.

In view of the predictive values found for the Strong Writer-Lawyer scale in this study, and the support found in other studies, it would seem that more extensive investigation of the scale and its independence from general ability measures should be attempted. It is suggested that this be done separately for each college, with as representative a group as can be obtained, and that it be combined with a scholastic aptitude test for determination of their multiple correlation with college grades.

While the nature of the data in this study permits no definite conclusions therefrom, certain speculations may be permitted.
It seems highly probable that the low coefficients of correlation with the criterion that were demonstrated for the test scores of the group in this study are in large part a function of the restricted range of the sample. This in turn is related to the counseling process--in that counselees with unfavorable patterns of scores are frequently discouraged from attempting the college curriculum. If the criterion scores of the 80 counselees who did not enter or re-enter the engineering course after counseling, and the 19 who withdrew or transferred soon after counseling, could have been included in the study, the coefficients of correlation obtained might have been markedly increased in magnitude.

It follows that the selection resulting from the counseling process reduces correlation coefficients by increasing the homogeneity of the group who attempt college after counseling.

As testing and counseling are extended downward into the high schools, and as the process becomes more widespread at college and pre-college levels, it can be expected to become more difficult to demonstrate satisfactory coefficients of correlation between the tests used by the counselors and the point-hour ratios obtained by those who go to college. It is therefore suggested that the validity of such tests be demonstrated by procedures other than correlation and that their predictive efficiencies be described by a different statistic.

- 32 -
A possible procedure might be to test the entire entering classes prior to entrance; and then to test the survivors at the beginning of their sophomore, junior, and senior years. Taking into consideration the question of what proportion of the unpredicted variance of criterion scores occurs in a positive direction, and the upper limits of magnitude of such deviations in individuals with low test scores—it should be practicable to set cutting scores on the tests below which only a very minute portion of the students would ever succeed. This procedure should be repeated at each class level in order to provide counselors with maximally useful data.

Such scores would serve the counselor far more adequately than correlation coefficients and multiple regression equations developed therefrom—since the counselor's usual problem is the dividing of students into a pass-fail dichotomy—and not that of predicting relative success.
BIBLIOGRAPHY


