Predicting achievement of first year technical students at North Iowa Area Community College

Thomas Eugene Ransford
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/rtd

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
Ransford, Thomas Eugene, "Predicting achievement of first year technical students at North Iowa Area Community College" (1967). Retrospective Theses and Dissertations. 16035.
http://lib.dr.iastate.edu/rtd/16035

This Thesis is brought to you for free and open access by Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
PREDICTING ACHIEVEMENT OF FIRST YEAR TECHNICAL STUDENTS AT NORTH IOWA AREA COMMUNITY COLLEGE

by

Thomas Eugene Ransford

A Thesis Submitted to the Graduate Faculty in Partial Fulfillment of The Requirements for the Degree of MASTER OF SCIENCE

Major Subject: Education

Signatures have been redacted for privacy

Iowa State University Of Science and Technology Ames, Iowa

1967
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>REVIEW OF LITERATURE</td>
<td>3</td>
</tr>
<tr>
<td>METHOD OF PROCEDURE</td>
<td>7</td>
</tr>
<tr>
<td>Basic Assumptions</td>
<td>7</td>
</tr>
<tr>
<td>Criterion of Success</td>
<td>8</td>
</tr>
<tr>
<td>Prediction Variables</td>
<td>8</td>
</tr>
<tr>
<td>1. High school mathematics average</td>
<td>8</td>
</tr>
<tr>
<td>2. Quantitative Thinking percentile rank of the Iowa Tests of Educational Development</td>
<td>9</td>
</tr>
<tr>
<td>3. Composite percentile rank of the Iowa Tests of Educational Development</td>
<td>9</td>
</tr>
<tr>
<td>4. The rank of the student in his high school class</td>
<td>10</td>
</tr>
<tr>
<td>Null Hypothesis</td>
<td>10</td>
</tr>
<tr>
<td>Statistical Technique</td>
<td>10</td>
</tr>
<tr>
<td>FINDINGS</td>
<td>13</td>
</tr>
<tr>
<td>Four-variable Predictor</td>
<td>16</td>
</tr>
<tr>
<td>Three-variable Predictor</td>
<td>17</td>
</tr>
<tr>
<td>Two-variable Predictor</td>
<td>20</td>
</tr>
<tr>
<td>One-variable Predictor</td>
<td>22</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>25</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>29</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>31</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. Mean scores and standard deviations of the criterion and variables 14
Table 2. Product-moment correlation matrix 14
Table 3. Summary of multiple regression 15
Table 4. Analysis of multiple regression using four variables 17
Table 5. Analysis of multiple regression using three variables 18
Table 6. Test for loss in prediction of the criterion due to the elimination of the ITED Composite from the four-variable regression equation 19
Table 7. Analysis of multiple regression using two variables 20
Table 8. Test for loss in prediction of the criterion due to the elimination of high school mathematics average from the three-variable regression equation 21
Table 9. Analysis of regression using one-variable 23
Table 10. Test for loss in prediction of the criterion due to the elimination of the ITED Quantitative Thinking from the two-variable regression equation 24
INTRODUCTION

The enrollment in the secondary schools has been increasing over the past years. At the same time, technological advancements require that high school graduates be more highly trained before entering the labor market. These two factors combine to create a larger enrollment in institutions offering post high school training. As this enrollment increases, facilities become crowded and it is important to select for admission those students who have the best chance of succeeding. This selection is also desirable to prevent students from entering a vocational field for which they are not suited. To accomplish this selection it is imperative that counseling and admissions personnel have reliable measures to predict the probable success of applicants.

The purposes of this study are two-fold. The first is to determine the predictive value of four variables in determining success at the end of the first year of the two year technical programs at North Iowa Area Community College. The second is to develop a regression formula from these variables by which first year achievement may be predicted.

The four variables which are to be considered for each student are as follows:

1. The average of semester grades from all mathematics courses taken while in high school.

2. The Quantitative Thinking percentile rank of the Iowa Tests of Educational Development.
3. The Composite percentile rank of the Iowa Tests of Educational Development.

4. The rank of the student in his high school class. While other scores are available for individual students, these are the four which will be consistently available in the future and which have been the most reliable in predicting academic success in other studies.

North Iowa Area Community College, formerly Mason City Junior College, awards a degree of Associate in Applied Science to all students who complete the prescribed curriculum with a minimum average of 2.0 on a 4.0 scale (G). There are six areas of specialization open to entering students in the technical programs: electronics, drafting, refrigeration and air conditioning, automotive, quantity food service, and agriculture and light equipment. The school also has two separate nursing programs which were not included in this study because of variance in organization, entrance requirements, and length of course. All curriculums are approved by the Vocational Division of the Iowa State Department of Public Instruction.
REVIEW OF LITERATURE

Many studies have been conducted to predict academic achievement of various types using a large range of predictive variables. Among these studies are numerous attempts at predicting success in technical fields. This review is concerned with studies of technical success using the same or similar variables as the ones to be used in this study. Also reviewed were studies using the same or similar variables to predict success in other than technical endeavors.

Bloom and Peters (1) indicated that, despite the variation caused by the error of teachers' judgements about the quality of students' academic achievement, this source of variation is not as great as had been thought. Grade averages may have a reliability as high as .85. They also found a corrected correlation between high school mathematics grades and success in post high school mathematics related courses of .59.

In a study of 16 institutions of varying types, size, and location, Frederiksen (4) found a correlation of from .43 to .68 between high school standing and first year college grades. The median was found to be .57 of all schools studied.

High school percentile rank was found to be the best single predictor of college marks when compared to the American Council on Education Psychological Examination and the Ohio State University Psychological Examination (r = .58) by Samenfeld (14). The multiple correlation of high school
percentile rank and the Ohio State University Psychological Examination was .68.

A study was made in 1955 by Kacalek (6) to determine the effectiveness of the Iowa Tests of Educational Development in predicting high school achievement.

"It was found that when one variable was used, a satisfactory prediction could be obtained by using the Composite Score on the Iowa Tests of Educational Development, as the coefficient of correlation for this variable was .6290. The Multiple correlation of .6326 indicated that a good two-variable prediction could be obtained from the Composite Score and Intelligence Quotient."

Kacalek summarized that it appeared possible to predict high school academic achievement with Intelligence Quotient and the scores of the Iowa Tests of Educational Development without the use of tests designed to forecast academic achievement.

Kalajer (7) used the Bennett Test of Mechanical Comprehension, the Minnesota Paper Form Board Test, a locally-constructed mechanics test, and the Iowa Tests of Educational Development to predict achievement in a high school mechanics course at Mason City High School. Singly, all gave correlations between .44 and .45. The locally-constructed test was dropped from the multiple regression without significant loss of predictive ability. A multiple R of .56 was found for the remaining three variables.

The Iowa Tests of Educational Development were also used by Trueblood (15) in predicting high school algebra success. He found that when one variable was used, the Composite Score
of the Iowa Tests of Educational Development yielded a coefficient of correlation of .6537. The multiple correlation of .6803 indicated that the best two-variable prediction could be obtained from the Intelligence Quotient and the Composite Score of the Iowa Tests of Educational Development.

High school average grades in five areas (English, foreign language, social studies, mathematics, and natural sciences) were used along with the American Council on Education Psychological Examination to predict success in the various divisions of Iowa State College by Cation (3) in 1939. An English placement test was also used, but was dropped from the multiple regression without significant loss. High school average grades for the areas mentioned yielded correlations with first quarter achievement of .55 in the Division of Engineering, .56 in the Division of Science, and .59 in all divisions together.

Killam (9) used eighth grade English, mathematics, general science, and social science to predict achievement in three industrial arts courses: general industries, woodshop, and mechanical drawing. In general industries the eighth grade mathematics marks alone were as good at prediction as were all variables combined. Eighth grade mathematics score and general science score combined predicted achievement in woodshop as well as a combination of all four variables. The mathematics score and social science score combined proved to be the best predictor in the mechanical drawing area.

In a similar study, Omwig (12) found a multiple R of .61
with eighth grade grade point average and the Differential Aptitude Test, mechanical reasoning section, for predicting achievement in first year high school machine shop. A correlation of .49 was found between the same two variables and first year achievement in woodshop. Omwig (12) concluded:

"The most consistent variables in both studies were those representing previous teachers' marks."

It was found in a study by Cain (2) concerning various research techniques for predicting academic achievement that:

"Many investigators report that high school marks generally provide a more accurate basis for the prediction of college scholarship than do intelligence tests."

"In general, the techniques found useful in educational prognosis have been found useful in vocational prognosis."

In conclusion, there is evidence from this review of literature that achievement in technical areas has been predicted with some measure of success. It is further indicated that the scores of the Iowa Tests of Educational Development and marks assigned by previous teachers are of value in making these predictions.
METHOD OF PROCEDURE

The purposes of this study were to determine the value of four variables in predicting first year success of technical students at North Iowa Area Community College and to develop a regression formula using some combination of one or more of these variables.

The sample used consisted of 155 students completing the first year of the technical programs at North Iowa Area Community College, hereafter referred to as NIACC, between the years of 1962 and 1967. Data were collected from the permanent records of the school and the students' cumulative folders. Only the records of male students for whom all variable scores were available were used. Students who transferred to NIACC from other schools were excluded. The data were coded, punched on IBM cards, verified, and analyzed on the 360 IBM Computer by the Iowa State University Computation Center. All possible combinations of the variables were compared to the criterion by the technique of analysis of regression.

Basic Assumptions

For the purposes of this study, the following basic assumptions were made:

1. A linear relationship existed between the criterion and the variables.

2. Cumulative grade point average at the end of the first year is a satisfactory indication of successful completion of the first year.
3. Grade differences caused by students having attended different high schools under varying instructors will introduce negligible bias.

Criterion of Success

The criterion of success for this study was the grade point average at the end of the first year of the technical programs. Letter grades were given the students in the various courses taken and were changed to a numerical value according to the following scale:

- 4 indicated an A
- 3 indicated a B
- 2 indicated a C
- 1 indicated a D
- 0 indicated an F

Each course carried a specified number of semester hours of credit. The hours of credit were multiplied by the number value of the grade received and an arithmetic average was computed of all courses taken during the first year for each student. The criterion will hereafter be referred to as the first year grade point average or Y value.

Prediction Variables

1. High school mathematics average

The first variable used was the average of the semester grades of all mathematics courses taken by the student while in high school. The letter grade was changed to a numerical value according to the scale used for the criterion above. Where
high school grades were recorded as percentages, they were first converted to letter grades by the systems of the individual high schools, and then to numerical values. A simple arithmetic average was then computed. This variable was referred to as the high school math average or $X_1$.

2. Quantitative Thinking percentile rank of the Iowa Tests of Educational Development

The Ability to do Quantitative Thinking, Test 4, of the Iowa Tests of Educational Development was described as a general mathematics test with problems of a practical nature (5). It was used in percentile rank form because this was the form which was most readily available. It was felt by officials of NIACC that this would be the form of this score which would be most readily available in the future. This variable was referred to as the ITED Quantitative Thinking or $X_2$.

3. Composite percentile rank of the Iowa Tests of Educational Development

The Composite Score of the Iowa Tests of Educational Development was used for the third variable. It was designed to give a fairly accurate indication of the general level of the pupil’s educational development. It was obtained by finding the sum of the standard scores on Tests 1-8 and changing this to a standard score by means of a predetermined table (5). It was used in percentile rank form for the same reasons as was variable $X_2$. This variable was referred to as the ITED Composite or $X_3$. 
4. The rank of the student in his high school class

The rank in class of high school students was expressed as the student's numerical position with relation to his classmates as determined by his average grade at the end of the senior year. The student with the highest average ranked number one, while the rank of the student with the lowest average was equal to the number of students in the graduating class. For the purpose of this study, the number of students in the class was divided into the individual student's class rank in order to obtain a comparative figure. This variable was referred to as high school class rank or $X_4$.

Null Hypothesis

A null hypothesis was assumed for the purpose of statistical inference. It was stated as follows: High school mathematics average, ITED Quantitative Thinking, ITED Composite, and high school class rank are of no significant value when used as predictors of achievement at the end of the first year of the technical programs at NIACC.

The tenability of this hypothesis depended upon the degree of significance found when the data were subjected to statistical treatment. A significant $F$ value was considered sufficient evidence to reject this hypothesis.

Statistical Technique

All possible combinations of the variables were computed by the regression program of the Iowa State University
Computation Center.

The regression equation (17, p. 237) in raw score form, used to predict the criterion with four variables was:

\[ Y = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + C \]

where

\[ Y = \text{achievement after first year at NIACC} \]
\[ X_1 = \text{high school math average} \]
\[ X_2 = \text{ITED Quantitative Thinking percentile rank} \]
\[ X_3 = \text{ITED Composite percentile rank} \]
\[ X_4 = \text{high school class rank} \]
\[ a_1, a_2, a_3, a_4, \text{ and } C = \text{appropriate constants whose values were obtained through the analysis of multiple regression} \]

In determining the usefulness of each variable for predicting achievement in the criterion, analysis of regression tables were computed for all possible combinations of the variables. The best combinations of variables were determined by selecting those with the highest coefficient of correlation, and the lowest standard error of estimate in each variable group of four, three, two, and one. The coefficient of correlation was an indicator of the proportion of total variance in the criterion value which was due to the prediction variables. A reduction of the amount of error in prediction was indicated by a low standard error of estimate.

In selecting a variable for prediction, a high correlation with the criterion and a low intercorrelation with the other
variables was desired. A high correlation with the criterion indicated a favorable relationship between that variable and the criterion. Low intercorrelation between variables indicated that the variables were measuring dissimilar characteristics.

Beginning with the four-variable prediction combination, variables were dropped one at a time and an F test was used to measure the level of significant loss. This process was continued until a significant loss indicated a marked decrease in predictive ability. The remaining variable or variables were then used to write the regression formula for predicting the criterion.
FINDINGS

The data were collected and sent to the Iowa State University Computation Center for processing. The results were further analyzed in evaluating the four variables used as predictors of the criterion. Mean scores and standard deviations are listed in Table 1 for the variables and the criterion.

Coefficients of correlation were computed between the criterion and the prediction variables. Table 2 represents a correlation matrix which lists also the intercorrelations between the individual prediction variables. The highest correlation with the criterion was found to be high school class rank (-0.43), followed by the ITED Composite (0.38), the ITED Quantitative Thinking (0.37), and high school math average (0.24). The negative value of the correlations associated with high school class rank were caused by the inverse order of class rank; the highest student ranked number one while the lowest ranked 100.

The summary of analysis of multiple regression, Table 3, contains computed F values, multiple correlations, and standard error of estimate values for each combination of the prediction variables. This table was used to select the best combination of variables in each group of four, three, two, and one. Selection was made on the basis of largest F value, highest multiple R, and the lowest standard error of estimate.
Table 1. Mean scores and standard deviations of the criterion and variables (N = 155)

<table>
<thead>
<tr>
<th>Criterion and variables</th>
<th>Mean score</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$ High school math average</td>
<td>2.07729032</td>
<td>0.677234934</td>
</tr>
<tr>
<td>$X_2$ ITED Quantitative Thinking</td>
<td>65.1096774</td>
<td>23.2402941</td>
</tr>
<tr>
<td>$X_3$ ITED Composite</td>
<td>56.3548387</td>
<td>24.1143763</td>
</tr>
<tr>
<td>$X_4$ High school class rank</td>
<td>57.1290323</td>
<td>20.7216273</td>
</tr>
<tr>
<td>$Y$ First year grade point average</td>
<td>2.21077419</td>
<td>0.615415112</td>
</tr>
</tbody>
</table>

Table 2. Product-moment correlation matrix

<table>
<thead>
<tr>
<th>Variables</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$ High school math average</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_2$ ITED Quantitative Thinking</td>
<td>0.4192</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_3$ ITED Composite</td>
<td>0.3161</td>
<td>0.7950</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_4$ High school class rank</td>
<td>-0.5668</td>
<td>-0.4019</td>
<td>-0.4761</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>$Y$ First year grade point average</td>
<td>0.2439</td>
<td>0.3724</td>
<td>0.3773</td>
<td>-0.4266</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Table 3. Summary of multiple regression

<table>
<thead>
<tr>
<th>Variables used for prediction No.</th>
<th>Variables eliminated No.</th>
<th>F values significant at 1% level</th>
<th>F values significant at 5% level</th>
<th>Ry</th>
<th>Standard error of estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>$X_1X_2X_3X_4$</td>
<td>11.50605605</td>
<td>-</td>
<td>.48454974</td>
<td>.545473486</td>
</tr>
<tr>
<td>3</td>
<td>$X_1X_2X_3$</td>
<td>10.07778686</td>
<td>-</td>
<td>.40843612</td>
<td>.567295639</td>
</tr>
<tr>
<td>3</td>
<td>$X_1X_2X_4$</td>
<td>15.30156811</td>
<td>-</td>
<td>.48283701</td>
<td>.544252548</td>
</tr>
<tr>
<td>3</td>
<td>$X_1X_3X_4$</td>
<td>14.30377859</td>
<td>-</td>
<td>.47041853</td>
<td>.548437217</td>
</tr>
<tr>
<td>3</td>
<td>$X_2X_3X_4$</td>
<td>15.26414859</td>
<td>-</td>
<td>.48238379</td>
<td>.544407758</td>
</tr>
<tr>
<td>2</td>
<td>$X_1X_2$</td>
<td>13.20734728</td>
<td>-</td>
<td>.38477557</td>
<td>.571759260</td>
</tr>
<tr>
<td>2</td>
<td>$X_1X_3$</td>
<td>14.43559856</td>
<td>-</td>
<td>.39952843</td>
<td>.567863311</td>
</tr>
<tr>
<td>2</td>
<td>$X_1X_4$</td>
<td>16.90437410</td>
<td>-</td>
<td>.42656132</td>
<td>.560267510</td>
</tr>
<tr>
<td>2</td>
<td>$X_2X_3$</td>
<td>14.11498909</td>
<td>-</td>
<td>.39576898</td>
<td>.568872681</td>
</tr>
<tr>
<td>2</td>
<td>$X_2X_4$</td>
<td>22.71670464</td>
<td>-</td>
<td>.47970842</td>
<td>.543523335</td>
</tr>
<tr>
<td>2</td>
<td>$X_3X_4$</td>
<td>21.58506724</td>
<td>-</td>
<td>.47031087</td>
<td>.546665714</td>
</tr>
<tr>
<td>1</td>
<td>$X_1$</td>
<td>9.67931077</td>
<td>-</td>
<td>.24392485</td>
<td>.598773202</td>
</tr>
<tr>
<td>1</td>
<td>$X_2$</td>
<td>24.63806040</td>
<td>-</td>
<td>.37242193</td>
<td>.573007763</td>
</tr>
<tr>
<td>1</td>
<td>$X_3$</td>
<td>25.39628269</td>
<td>-</td>
<td>.37730467</td>
<td>.571788764</td>
</tr>
<tr>
<td>1</td>
<td>$X_4$</td>
<td>34.02963598</td>
<td>-</td>
<td>.42655343</td>
<td>.558435865</td>
</tr>
</tbody>
</table>

*a*Indicated the best combination of variables in each group.
Four-variable Predictor

The regression equation, in raw score form used to predict the criterion with four variables was:

\[ Y = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + C \]

where

- \( Y \) = First year grade point average
- \( X_1 \) = High school math average
- \( X_2 \) = ITED Quantitative Thinking percentile rank
- \( X_3 \) = ITED Composite percentile rank
- \( X_4 \) = High school class rank

The values of the regression constants obtained from the data for the four-variable prediction equation were as follows:

\( a_1 = -0.053233398 \)
\( a_2 = 0.00534148938 \)
\( a_3 = 0.0182052640 \)
\( a_4 = -0.0102377879 \)
\( C = 2.45585155 \)

The four-variable regression equation became:

\[ Y = -0.053233398X_1 + 0.00534148938X_2 + 0.0182052640X_3 \\
- 0.0102377879X_4 + 2.45585155 \]

The results of the four-variable multiple regression, Table 4, yielded a multiple correlation of 0.48. The F value of 11.51 was found to be highly significant beyond the one percent level. The null hypothesis was rejected.
Table 4. Analysis of multiple regression using four variables 
($X_1, X_2, X_3, X_4$)

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Sum of squares</th>
<th>Mean square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>4</td>
<td>13.6941086</td>
<td>3.42352714</td>
<td>11.50605605**a</td>
</tr>
<tr>
<td>Residual</td>
<td>150</td>
<td>44.6311385</td>
<td>0.297541323</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td>58.3253071</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard error = 0.55

$F_{4, 150} = 11.51**$

$R_y (1, 2, 3, 4) = 0.48$

Tabled values

| .01 | 3.44 |
| .05 | 2.43 |

*aTwo asterisks (**) were used to denote significance beyond the 1% level throughout the study.*

Three-variable Predictor

The best combination of three variables, shown in Table 3, was high school math average, ITED Quantitative Thinking, and high school class rank. Multiple regression for these variables was computed and is shown in Table 5. The analysis yielded a highly significant $F$ of 15.30, which indicated that this combination of variables could be used effectively in predicting the criterion. The multiple correlation was 0.48.

The constants for the three-variable regression equation became:
\[ a_1 = -0.062577775 \]
\[ a_2 = 0.00676347685 \]
\[ a_4 = -0.0107789248 \]
\[ G = 2.51617050 \]

The regression formula became:
\[
Y = -0.062577775X_1 + 0.00676347685X_2 - 0.0107789248X_4 + 2.51617050
\]

Table 5. Analysis of multiple regression using three variables 
\((X_1, X_2, X_4)\)

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Sum of squares</th>
<th>Mean square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>3</td>
<td>13.5974709</td>
<td>4.53249028</td>
<td>15.30156811**</td>
</tr>
<tr>
<td>Residual</td>
<td>151</td>
<td>44.7278362</td>
<td>0.296210836</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td>58.3253071</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard error = 0.54

\[ F_{3,151} = 15.30^{**} \]

Ry (1, 2, 4) = 0.48

Tabled values .01 3.91

.05 2.67
A test was computed to determine whether there was significant loss in predictive ability due to the elimination of the ITED Composite, X₃. This is shown in Table 6. The calculated F value of 0.32 was not significant, which indicated that the dropping of the X₃ variable would not affect the predictive ability of the regression equation.

Table 6. Test for loss in prediction of the criterion due to the elimination of the ITED Composite from the four-variable regression equation

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Sum of squares</th>
<th>Mean square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression (4 var.)</td>
<td>4</td>
<td>13.6941086</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression (3 var.)</td>
<td>3</td>
<td>13.5974709</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss (X₃)</td>
<td>1</td>
<td>0.0966377</td>
<td>0.0966377</td>
<td>0.3247874</td>
</tr>
<tr>
<td>Residual (4 var.)</td>
<td>150</td>
<td>44.6311985</td>
<td>0.2975413</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td>58.3253071</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F₁,150 = 0.32  
Tabled values .01 6.81 .05 3.91
Two-variable Predictor

The best two-variable predictor from Table 3 was composed of the ITED Quantitative Thinking and high school class rank. When these two variables were used in combination, the following constants were obtained:

\[
\begin{align*}
a_2 &= 0.00634729754 \\
a_4 &= -0.00980693460 \\
C &= 2.35776424 \\
\end{align*}
\]

The regression equation became:

\[
Y = 0.00634729754X_2 - 0.00980693460X_4 + 2.35776424
\]

The two-variable multiple regression using ITED Quantitative Thinking and high school class rank was computed (Table 7). The F value of 22.72 was highly significant indicating the criterion could be predicted using these two variables. The multiple coefficient of correlation was found to be 0.48.

Table 7. Analysis of multiple regression using two variables \( (X_2, X_4) \)

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Sum of squares</th>
<th>Mean square</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>2</td>
<td>13.4218295</td>
<td>6.71091473</td>
<td>22.71670464**</td>
</tr>
<tr>
<td>Residual</td>
<td>152</td>
<td>44.9034776</td>
<td>0.295417616</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td>58.3253071</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td></td>
<td>0.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ry (2,4)</td>
<td></td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F ( 2,152 )</td>
<td></td>
<td>22.72**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tabled values</td>
<td>.01</td>
<td>4.75</td>
<td></td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>5.06</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A test was computed to determine whether there was significant loss in predictive ability due to the elimination of the high school math average, $X_1$. The $F$ value of 0.59 was not significant (Table 8). This indicated that the criterion of first year grade point average could be predicted equally as well with the two-variable regression as with the addition of the high school math average variable.

Table 8. Test for loss in prediction of the criterion due to the elimination of high school mathematics average from the three-variable regression equation

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Sum of squares</th>
<th>Mean square</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression (3 var.)</td>
<td>3</td>
<td>13.5974709</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression (2 var.)</td>
<td>2</td>
<td>13.4218295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss ($X_1$)</td>
<td>1</td>
<td>0.1756414</td>
<td>0.1756414</td>
<td>0.592960</td>
</tr>
<tr>
<td>Residual (3 var.)</td>
<td>151</td>
<td>44.7278362</td>
<td>0.2962108</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td>58.3253071</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$F_{1,151} = 0.59$  
Tabled values .01 6.81 .05 3.91
One-variable Predictor

The best single-variable predictor shown in Table 3 was the high school class rank. This variable yielded constants as follows:

\[ a_4 = -0.0126682824 \]
\[ C = 2.93450069 \]

The regression equation became:

\[ Y = -0.0126682824X_4 + 2.93450069 \]

The computation of the single-variable regression using high school class rank, \( X_4 \), is shown in Table 9. The \( F \) value computed, 34.03, was significant beyond the 1% level, with a correlation of 0.43.

A test was computed to determine the loss in predictive ability due to the elimination of the ITED Quantitative Thinking variable, \( X_2 \). This test yielded an \( F \) value of 0.96, which was not significant (Table 10). This indicated that prediction of the criterion was not significantly affected by the elimination of the ITED Quantitative Thinking variable, and that it could be predicted as well with the single variable, high school class rank.

From the preceding analysis of data, it was found that the high school class rank was as effective in predicting the criterion of first year grade point average as any combination using this variable with the other three; high school math average, ITED Quantitative Thinking, and ITED Composite.

The single-variable equation which was a result of this
analysis was as follows:

\[ Y = -0.0126682824x_{4} + 2.03450069 \]

This equation represented the most satisfactory method of predicting the criterion with the variables used for this analysis of regression.

Table 9. Analysis of regression using one-variable (X4)

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Sum of squares</th>
<th>Mean square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>10.6121629</td>
<td>10.6121629</td>
<td>34.02963598**</td>
</tr>
<tr>
<td>Residual</td>
<td>153</td>
<td>47.7131442</td>
<td>0.311850615</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td>58.3253071</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard error = 0.56

\[ F_{1,153} = 34.03^{**} \]

Ry (4) = 0.43

Tabled values .01 6.81 .05 3.91
Table 10. Test for loss in prediction of the criterion due to the elimination of the ITED Quantitative Thinking from the two-variable regression equation

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Sum of squares</th>
<th>Mean square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression (2 var.)</td>
<td>2</td>
<td>13.4218295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression (1 var.)</td>
<td>1</td>
<td>10.6121629</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss ($X_2$)</td>
<td>1</td>
<td>2.80996666</td>
<td>2.80996666</td>
<td>.9573401</td>
</tr>
<tr>
<td>Residual (2 var.)</td>
<td>152</td>
<td>44.9034776</td>
<td>2.9348678</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td>58.3253971</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$F_{1,152} = 0.96$ Tabled values .01 6.81  .05 3.91
DISCUSSION

This study was conducted in an effort to find a satisfactory predictor of first year achievement of students enrolled in the two year technical program at North Iowa Area Community College. A high intercorrelation was found between the two variables associated with the Iowa Tests of Educational Development. This would indicate that they measure similar characteristics, which might have been expected, as the Quantitative Thinking score is one of several used to determine the Composite score. Because of the level of this intercorrelation, only one of these two variables would have been necessary for this study.

The ITED Composite was the first to be dropped from the four-variable regression with an insignificant $F$ value of 0.32. The high school math average and the ITED Quantitative Thinking variables were dropped in that order with $F$ values of 0.59 and 0.96, respectively. This left only the variable of high school class rank with a coefficient of correlation of -0.4266. The negative character of this coefficient is explained by the highest student in a class being ranked number one and the lowest being ranked number 100.

A correlation of this value, although significant, would be considered high for predicting achievement of a group but is not considered high for the purpose of predicting academic achievement of an individual (17, p. 76).
A review of the literature indicated that classroom teacher-assigned marks, and particularly rank in graduating class have been reasonably good predictors of academic achievement. In this study, class rank was found to be the best predictor. There are several factors which may have contributed to its not being as highly correlated to the criterion as in other studies of similar nature. First of these factors may have been the age and maturity of a portion of the cases included in the study. Many had been out of high school for a year or more before entering the technical program. Several had served their military obligation between the time they graduated from high school and the time they enrolled at NIACC.

Another of the contributing factors may have been the motivation of the students. While they may not have performed well in general education courses, the very nature of a technical program and its obvious applications may have caused a student to do better work in courses of this type.

A third contributing factor to the low coefficient of correlation may well have been the large range of technical programs included in this study. While it was desirable to have a prognostic indication of the success of students before admitting them to the program, this study may have been premature. At this point in the development of the area community college, there had been insufficient enrollment at NIACC to justify a predictive study in each of the six areas covered in this study. For this reason, the areas were treated
collectively. While some of the areas are closely related, there would seem to be a large differential in the level and type of abilities of a successful student of electronics technology and one who was a success in quantity food preparation, for example.

Further extenuating circumstances were related to the population of the study. The population consisted of all students who had been admitted to NIACC during the years considered. These students had already undergone some form of selection, either directly or indirectly. Some students coming from the high schools enrolled in four year college programs. Others did not attempt any type of post-high school training. Of those who did select the NIACC technical programs, not all were admitted. The result of this voluntary and involuntary selection was that the population consisted predominantly of middle- and lower-ability students.

The regression formula, page 30, would indicate that no student, even an individual ranking number one in his class, could achieve beyond the level of 2.02. While this was the result of the statistical analysis, this formula should, by no means, be used to determine whether an individual student would be selected for admission. The population was definitely skewed in favor of students with high class ranks. The sample contained only two students whose class rank was lower than 20 and the number of students in the higher deciles of class rank was not balanced by a proportional number in the lower deciles.
With such a sample, no attempt should be made to predict the achievement of individuals with this formula alone.

It is for the above reasons that, after a period of perhaps five years, a predictive study should be made of each of the technical areas separately. At that time a sufficient number of students should have completed the first and second years to warrant such a series of studies.
SUMMARY

The purposes of this study were to determine the value of four variables in predicting first year success of technical students at North Iowa Area Community College and to write a regression formula using some combination of one or more of these variables. Data were collected on 155 students who had completed the first year of one of six technical areas between the years of 1962 and 1967. An analysis of regression was computed by a 360 IBM Computer using these data. The variables were then dropped one at a time from the regression formula and a test was computed to determine the loss of prediction ability associated with this elimination. The criterion of success in each case was first year grade point average.

The four-variable analysis of regression yielded a correlation of 0.48 with a highly significant F value of 11.51. First of the variables to be dropped was the ITED Composite. The resultant loss was insignificant as indicated by a 0.32 F value. The three-variable analysis of regression was then computed with a correlation of 0.48. The F value of 15.30 was significant beyond the one percent level.

Subsequently, the variable of high school math average was dropped from the three-variable regression and an F value of 0.59 indicated that there was no significant loss associated with the elimination of this variable. The remaining two-variable analysis of regression showed a multiple coefficient of
correlation of 0.48. The F value of 22.72 was significant beyond the one percent level.

The variable of the ITED Quantitative Thinking was then dropped from the two-variable regression. Here again the loss was insignificant as indicated by an F value of 0.96. The remaining single variable of high school class rank showed a coefficient of correlation with the criterion of first year grade point average of -0.43. The F value of 34.03 was significant beyond the one percent level. The negative coefficient was explained by the inverse arrangement of high school class rank.

This single variable was used with its associated constants to write a regression formula which appeared as follows:

\[ Y = -0.0126682824X_4 + 2.03450069 \]

This equation represents the most satisfactory method of predicting the criterion of first year grade point average with the best combination of the variables of high school math average, ITED Quantitative Thinking percentile rank, ITED Composite percentile rank, and high school class rank.

Table II represents the level of achievement of first year students by deciles of high school class rank (X_4) as predicted by this regression equation. However, when using this table, the level of the coefficient of correlation (-0.43) and the limitations of statistical predictions should be kept in mind (see DISCUSSION, page 25).
BIBLIOGRAPHY


5. Iowa Tests of Educational Development fall testing program: how to use the test results; a manual for teachers and counselors. Ninth edition. [Iowa City, Iowa, Office of Iowa Testing Programs, University of Iowa]. 1964.


