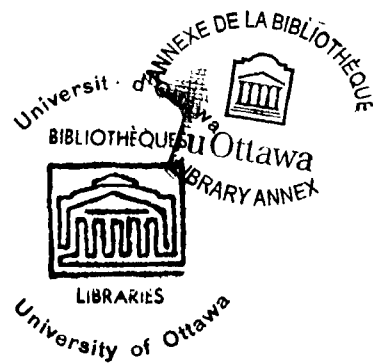


CSP-2  
001585

THE RELATIVE CONTRIBUTION OF SOME  
DIVERGENT THINKING TESTS TO THE  
PREDICTION OF ACADEMIC SUCCESS

by W.K. Noë

Thesis presented to the Faculty of  
Psychology of the University of  
Ottawa as partial fulfillment of the  
requirements for the degree of  
Master of Arts in Psychology



Ottawa, Canada, 1971

UMI Number: EC55389

### INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI<sup>®</sup>

---

UMI Microform EC55389  
Copyright 2011 by ProQuest LLC  
All rights reserved. This microform edition is protected against  
unauthorized copying under Title 17, United States Code.

---

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

## ACKNOWLEDGEMENTS

This thesis was prepared under the supervision of Dr. Laurent A. Isabelle. The writer is grateful for his interest and counsel.

The writer is also indebted to Dr. O. Porebski for his advice on statistical matters, to Miss J. Radford for her assistance in the organizational aspects of the study, and to Mr. J.A.C. Préfontaine for the use of computer services for this project.

## CURRICULUM STUDIORUM

Walter K. Noë was born February 7, 1940 in Montréal, Québec. He received his Bachelor of Arts degree from Loyola of Montréal in 1960.

## TABLE OF CONTENTS

Chapter	page
INTRODUCTION . . . . .	vi
I.- REVIEW OF THE LITERATURE . . . . .	1
1. Prediction of Academic Success by IQ and Aptitude Tests	1
2. Prediction of Academic Success by Divergent Thinking Tests	3
a. Extreme Group Studies	5
b. Multiple Correlation Studies	14
c. The Threshold Hypothesis	20
3. The Problem and Hypothesis	20
II.- EXPERIMENTAL DESIGN . . . . .	23
1. The Sample	23
2. The Measuring Instruments	24
a. Criterion Measures	24
b. Predictor Measures	25
3. The Data Gathering Procedures	27
4. Reliability	28
5. Statistical Analysis	34
6. Subsequent Analyses	38
7. Statistical Assumptions	39
III.- RESULTS AND DISCUSSION . . . . .	41
1. Description of the Data	41
2. Validation and Cross-Validation Results	41
3. Summary and Discussion	54
BIBLIOGRAPHY . . . . .	56
 Appendix	
1. TESTS FOR LINEARITY OF REGRESSION . . . . .	59
2. FREQUENCY DISTRIBUTIONS FOR EACH VARIABLE AND SOME ASSOCIATED CHI SQUARE TESTS FOR GOODNESS OF FIT . . . . .	64
3. ABSTRACT OF <u>The Relative Contribution of Some Divergent Thinking Tests to the Prediction of Academic Success</u> . . . . .	78

## LIST OF TABLES

Table	page
I.- Split-Half Reliability Coefficients of the DAT Subtests on a Grade 12 Population . . . .	29
II.- Reliabilities of the DT Tests . . . . .	31
III.- Intercorrelations Among Parts of the DT Tests .	32
IV.- Means and Standard Deviations of the Parts of the DT Tests . . . . .	33
V.- Interscorer Reliability on the DT Tests . . . .	35
VI.- Correlations Between Mid-Year and Matriculation Scores . . . . .	36
VII.- Data Description of the Validation Sample . . .	42
VIII.- Data Description of the Cross-Validation Sample	43
IX.- Intercorrelations Among All the Tests on the Validation and Cross-Validation Groups . . . .	44
X.- Multiple R's and Cross-Validation R's . . . . .	46
XI.- Intercorrelations Among All the Tests on the Calibration and Verification Groups . . . . .	50
XII.- Multiple R's and Cross-Validation R's, Verifi- cation Group Predicted From the Calibration Group . . . . .	51
XIII.- Multiple R's and Cross-Validation R's, Cali- bration Group Predicted From the Verification Group . . . . .	52
XIV.- Regression of the GRA Criterion on the Nine Predictors . . . . .	59
XV.- Regression of the LIT Criterion on the Nine Predictors . . . . .	60
XVI.- Regression of the WRI Criterion on the Nine Predictors . . . . .	61

LIST OF TABLES

Table	page
XVII.- Regression of the ORA Criterion on the Nine Predictors . . . . .	62
XVIII.- Regression of the HIS Criterion on the Nine Predictors . . . . .	63
XIX.- NA Chi Square Test . . . . .	64
XX.- HIS Chi Square Test . . . . .	65
XXI.- VR Chi Square Test . . . . .	66
XXII.- AR Chi Square Test . . . . .	67
XXIII.- DMS Chi Square Test . . . . .	68
XXIV.- OTIS Frequency Distribution . . . . .	69
XXV.- DSU Frequency Distribution . . . . .	70
XXVI.- DMC Frequency Distribution . . . . .	71
XXVII.- DFT Frequency Distribution . . . . .	72
XXVIII.- DMR Frequency Distribution . . . . .	73
XXIX.- GRA Frequency Distribution . . . . .	74
XXX.- LIT Frequency Distribution . . . . .	75
XXXI.- WRI Frequency Distribution . . . . .	76
XXXII.- ORA Frequency Distribution . . . . .	77

## INTRODUCTION

Tests of cognitive abilities have long been used in the prediction of academic success. However, predictive efficiency seems to have reached a ceiling represented by a multiple correlation of approximately .6. Related to this is the often noted observation that multiple correlations seldom increase significantly after four or five tests have been inserted into the equation. Guilford suggests that this occurs because only a limited number of the numerous possible human abilities that are involved in a practical criterion have been represented in the tests used. Although a large number of different tests may have been tried out, the same limited number of fundamental factors have been measured by them.

Tests developed from Guilford's Structure-of-Intellect appear to offer new factors of intellectual ability. This is particularly true of the divergent production abilities which are "... conspicuous by their absence in modern group tests of intelligence ..."<sup>1</sup> These tests have been identified through factor analysis and can be said to have construct, or internal validity. However, little investigation has been carried out regarding their relation to external criteria of success.

---

<sup>1</sup> J.P. Guilford, The Nature of Human Intelligence, New York, McGraw-Hill, 1967, p. 138.

The purpose of this study is to investigate the incremental validity of a battery of divergent thinking tests when added to a standard aptitude battery in the prediction of academic success. An attempt is made thereby both to increase the level of academic prediction and to investigate the external validity of the divergent thinking tests.

Chapter I reviews the literature on the relationship of cognitive tests to academic success, concluding with the null hypothesis of this study. Chapter II presents the experimental design of the study while Chapter III presents the results and a discussion.

## CHAPTER I

### REVIEW OF THE LITERATURE

This chapter reviews the literature on the prediction of academic success by means of cognitive tests. First, the predictive efficiency of intelligence and aptitude tests is discussed. Since this is a well known area, having changed little for 35 years<sup>1</sup>, it will not be dealt with in detail.

Secondly, the relationship of divergent thinking (DT) tests to academic success is extensively reported. Finally, evidence for the hypothesis that DT tests account for criterion variance only after an IQ threshold has been reached is explored. The chapter concludes with the Null Hypothesis which will be tested in this study.

#### 1. Prediction of Academic Success by IQ and Aptitude Tests

In this study, the terms IQ and aptitude will be used in the usual sense. The term standard aptitude test will be used in a generic sense to refer to all cognitive tests which are not Divergent Thinking tests as this is the usual distinction made in the literature.

---

<sup>1</sup> W.B. Michael, "A Short Evaluation of the Research Reviewed in Educational and Psychological Testing", Review of Educational Research, Vol. 35, No. 1, 1965, p. 95.

Michael has commented, "In the past 30 years relatively little gain has occurred in the observed magnitude of validity coefficients of tests or batteries employed in the prediction of success in the academic setting ..."<sup>2</sup>. In elementary and high school, the average observed magnitude of the validity coefficients of intelligence tests is from .4 to .6 according to Tyler<sup>3</sup>, .5 according to Super and Crites<sup>4</sup> and .6 according to Thorndike and Hagen<sup>5</sup>.

Standard aptitude tests fare no better. Typically, verbal and numerical ability tests which share much common variance with intelligence tests are the best predictors of individual subjects or grade point average (GPA), and predictions are of the same level of magnitude as those achieved by intelligence tests<sup>6</sup>.

McNemar<sup>7</sup> in reviewing the validity studies reported in the Manual for the Differential Aptitude Tests counted

---

2 Idem, ibid., p. 95.

3 Leona E. Tyler, The Psychology of Human Differences, 3rd Edition, New York, Appleton Century Crofts, 1965, p. 112.

4 Donald E. Super and John O. Crites, Appraising Vocational Fitness by Means of Psychological Tests, New York, Harper, 1962, p. 87.

5 Robert L. Thorndike and Elizabeth Hagen, Measurement and Evaluation in Psychology and Education, Third Edition, New York, John Wiley and Sons, 1969, p. 324.

6 Quinn McNemar, "Lost: Our Intelligence? Why?", American Psychologist, Vol. 19, No. 12, 1964, p. 875.

7 Idem, ibid., p. 874.

4096 validity coefficients ranging from  $-.37$  to  $.90$ . Obviously, judicious selection of validity coefficients could show that the Differential Aptitude Tests (DAT) were either useless or unexcelled for academic prediction. From a careful appraisal of this data, McNemar<sup>8</sup> concluded that the most useful academic predictors were the Verbal Reasoning, Language Usage, Numerical Ability and Spelling tests in that order. Of these, the Language Usage and Spelling tests are more nearly achievement than aptitude tests. Isabelle<sup>9</sup>, in 1961 and Lavin in 1965<sup>10</sup> have presented detailed reviews of the efficacy of standard aptitude tests in the prediction of academic success.

## 2. Prediction of Academic Success by Divergent Thinking Tests

Two issues must be commented upon. First, since Guilford<sup>11</sup> has pointed out that DT tests are logically related to the hypothesized abilities underlying creative thinking, most authors refer to them as creativity tests.

---

8 Idem, ibid., p. 875.

9 Laurent Isabelle, Actuarial Versus Clinical Methods in Predicting Achievement in a Science Faculty, PhD Thesis presented to the School of Psychology and Education of the University of Ottawa, Ontario, 1961, xiv-176 p.

10 David E. Lavin, The Prediction of Academic Performance, New York, John Wiley and Sons, 1965, p. 47-63.

11 J.P. Guilford, "Creative Abilities in the Arts", in Research in Personality, Martha T. Mednick and Sarnoff A. Mednick, Editors, New York, Holt, Rinehart and Winston, 1963, p. 576.

Since the issue of creativity is irrelevant to this review, whenever a study cited uses the term creativity, it will be translated into divergent thinking.

A second issue is the multiplicity of DT measures used. A number of authors have made minor to major changes in the Guilford tests. Chief among these is Torrance<sup>12</sup> who has constructed the Minnesota Tests of Creative Thinking (MTCT). Originally, these were adaptations of Guilford's tests with a different scoring system<sup>13</sup>. Whereas the majority of Guilford's tests yield only one score, Torrance scores each test for fluency (number of responses), flexibility (shifts in the classes of answers), originality (statistical infrequency) and elaboration (number of details). These correlated scores are summed into part scores according to whether the stimulus material is verbal or non-verbal. The part scores are then summed into a total score. More recently, the trend in Torrance's thinking has been towards constructing tests which are models of the creative thinking process. Torrance claims that these tests represent a fairly sharp departure from the Guilford tests<sup>14</sup>. Since the

---

12 E. Paul Torrance, Guiding Creative Talent, Englewood Cliffs, Prentice-Hall, 1962, xi-278 p.

13 Idem, ibid., p. 44-47.

14 E. Paul Torrance, "Examples and Rationales of Test Tasks for Assessing Creative Abilities", Journal of Creative Behavior, Vol. 2, No. 3, 1968, p. 167.

majority of research into academic prediction has been done with Torrance's tests, those results will be reported in the following review. However, it is suggested that the aforementioned differences in theory and scoring may account for some of the discrepancies among the various studies. If the tests used in the cited studies were from neither the Guilford nor Torrance batteries, their contents are described.

This review will examine the contribution of DT tests to academic prediction in primary and secondary schools in three stages: (a) extreme group studies; (b) multiple correlational studies; (c) a threshold study.

#### a. Extreme Group Studies

Prototypical of the extreme group studies is the research reported by Getzels and Jackson<sup>15</sup>. Five Guilford type DT tests were given to 449 students in a laboratory school. Two extreme groups were identified. The first group consisted of those 28 students who were in the top 20% on IQ but below the top 20% on the combined DT measures (hereinafter called the Hi-Low group). The second group consisted of those 26 students who were in the top 20% on the combined DT measures but below the top 20% in IQ (hereafter called the Low-Hi group). The IQ scores were either the

---

<sup>15</sup> Jacob W. Getzels and Philip W. Jackson, Creativity and Intelligence: Explorations with Gifted Students, New York, John Wiley and Sons, 1962, p. 13-32.

Stanford-Binet or regression estimated Binet IQ's derived from either the Wechsler Intelligence Scale for Children (WISC) or the Henmon-Nelson Test. The educational level of the sample ranged from the 6th to the 12th grade. The achievement measures used were average scores derived from various standardized achievement tests.

The results showed that, on the achievement measures, the two experimental groups were equally superior to the rest of the population and not significantly different from each other.

The design of this study has been extensively criticized by De Mille and Merrifield<sup>16</sup>, McNemar<sup>17</sup> and Marsh<sup>18</sup>. Many of these criticisms can be applied to most of the studies using this type of extreme group method. The population was highly selected on the IQ variable and only incidentally selected on the DT measures, since the mean IQ of the total population was 132, that of the Hi-Low group 150 and

---

16 Richard De Mille and Philip R. Merrifield, Review of Creativity and Intelligence: Explorations with Gifted Students, by Jacob W. Getzels and Philip W. Jackson, in Educational and Psychological Measurement, Vol. 22, No. 4, 1962, p. 803-808.

17 McNemar, op. cit., p. 878-879.

18 R.W. Marsh, "Research Note: A Statistical Re-Analysis of Getzels and Jackson's Data", British Journal of Educational Psychology, Vol. 34, No. 1, 1964, p. 91-93.

that of the Low-Hi group 127. IQ scores are corrected for age, whereas DT scores are not. The DT scores were summed into a composite although their median intercorrelation of .28 was in effect no higher than their median correlation with the IQ measure of .26.

Finally, De Mille and Merrifield<sup>19</sup> point out that the practice of contrasting Hi-Low and Low-Hi groups is not comparable to the standard design which takes groups of extreme scorers, high and low, on the same measure. Instead, they say, "It is rather more like studying two types of people, parents and married people, while leaving out the group composed of married parents"<sup>20</sup>. The group high on both measures was ignored, as was the great majority of the sample population, i.e., those not in the top 20% on either measure. Hence, this type of design can reveal very little about the relative efficacy of DT tests as opposed to standard aptitude tests in predicting academic success.

Nevertheless, eight partial replications of the Getzels and Jackson study were conducted by Torrance<sup>21</sup>. The grade levels of the sample ranged from elementary school to university. The standard aptitude measures used varied from

---

19 De Mille and Merrifield, op. cit., p. 804.

20 Idem, ibid., p. 804.

21 Torrance, Guiding Creative Talent, op. cit., p. 54-66.

school to school. These included the Stanford-Binet, the Otis SA, the Kuhlmann-Anderson, the California Tests of Mental Maturity (CTMM), the Lorge-Thorndike tests, and the Miller Analogies Test. The DT measures used were the MTCT. The criterion measures were the Iowa Basic Skills Battery in five instances, the Iowa Tests of Educational Development in one instance and a specially constructed achievement test in two instances. The N's in each extreme group ranged from 10 to 36. The mean IQ's for the Hi-Low groups ranged from 114 to 152 and those for the Low-Hi groups ranged from 98 to 126.

In six of the eight groups no significant differences in achievement were found between the Hi-Low and Low-Hi groups. In the remaining two cases, the Hi-Low groups were significantly superior. Torrance<sup>22</sup> accounts for these two instances by noting that in one case the educational unit was a parochial school and in the other case it was a small town school. In both places, traditional (i.e., recall oriented) academic achievement was stressed. Thus, divergent thinking abilities apparently would not be utilized in such schools. Additionally, the two discrepant schools had more normal mean IQ's (98 and 107 for the Low-Hi group) than the other schools, suggesting to Torrance<sup>23</sup> that there is an IQ

---

22 Idem, ibid., p. 62.

23 Idem, ibid., p. 62.

threshold, only above which will DT tests begin to account for some unique variance in a criterion.

Torrance's results are difficult to interpret with any confidence because of the extreme group design, the diversity of criterion measures used and the selectivity on the IQ measure in some of the sample populations.

Another extreme group study which found the Getzels-Jackson phenomenon was conducted by Yamamoto<sup>24</sup> who sampled approximately 90% of the University of Minnesota High School. This is a laboratory school, comprising grades 9-12 where the average IQ on the Lorge-Thorndike Intelligence Test was 120. The Hi-Low and Low-Hi groups were identified in the usual manner. In addition, a Hi-Hi group, comprised of students who were in the top 20% on both DT and IQ measures was identified. The DT measures used were the Ask and Guess Test and the Test of Imagination from the MTCT battery. The criterion measures were the 8 Iowa Tests of Educational Development, expressed as standard scores. A summated total score was also used. The Hi-Low and Low-Hi groups had an N of 26 and the Hi-Hi group an N of 28. The groups were also divided by sex.

---

24 Kaoru Yamamoto, "Role of Creative Thinking and Intelligence in High School Achievement", Psychological Reports, Vol. 14, 1964, p. 783-789.

An Analysis of Variance showed no significant main effects or interactions. That is, subjects scoring high on divergent thinking but relatively low on IQ were just as successful academically as the high IQ groups.

Yamamoto<sup>25</sup> further divided his population into high and low DT groups (i.e., top 20% and bottom 20% on the composite DT measure). There were 54 subjects in each group. He found significant differences in favour of the high DT group between the achievement scores of those two groups when IQ was controlled by Analysis of Covariance. This lent further support to the hypothesis that high scores on DT tests are associated with academic success.

A number of studies have not found the Getzels-Jackson phenomenon.

Flescher<sup>26</sup> sampled the 6th grade population of a school district in New York State. The DT measures consisted of 3 Guilford-type tests plus a test of remote associations and of deviation from usual detail in a drawing. These were summed into a total score. The IQ measure was the CTMM

---

25 Kaoru Yamamoto, "A Further Analysis of the Role of Creative Thinking in High School Achievement", Journal of Psychology, Vol. 58, First Half, 1964, p. 277-283

26 Irwin Flescher, "Anxiety and Achievement of Intellectually Gifted and Creatively Gifted Children", Journal of Psychology, Vol. 56, Second Half, 1963, p. 251-268.

while the criterion measures were the 10 subscores and the total score of the Metropolitan Achievement Tests.

Four extreme groups were constructed. The Hi-Low group consisted of those subjects who had an IQ of 130 or above but who were not in the top quartile on DT (N = 28). The Low-Hi group consisted of those subjects who were in the top quartile on DT but whose IQ scores fell below 130 (N = 29). A third group, the Hi-Hi's consisted of those subjects whose DT scores were in the top quartile and whose IQ scores were 130 or higher (N = 24). A fourth group was selected from those whose IQ's were below 130 and whose DT scores fell below the top quartile (the Low-Low group, N = 29). For this latter group a subject was selected whose IQ score closely matched one IQ score in the Low-Hi group and who simultaneously possessed a DT score which closely approximated the DT measure of one individual in the Hi-Low group. The resulting group of 29 was characterized by means and SD's for IQ and DT which were of the same magnitude as their corresponding non-exceptional counterparts in the two otherwise gifted groups.

The mean IQ scores for the groups were: Hi-Low, 140; Low-Hi, 115; Hi-Hi, 142; Low-Low, 115.

Separate analyses of variance were carried out on each of the 10 subscores of achievement and the total score. In all cases, the main effect for IQ was significant beyond

the .01 level whereas neither the DT effect nor the interaction was significant. That is, IQ was related to academic success but DT scores were not.

Hasan and Butcher<sup>27</sup> conducted an experiment with a population unselected for IQ. They tested all 175 pupils in their second year of secondary education at a Scottish local authority comprehensive school. The IQ measure used was the Moray House Verbal Reasoning Test (Mean VRQ = 102, SD = 12). Ten Guilford-type DT measures were used. These were summed into a total score. The achievement measures were the Moray House English and Arithmetic attainment tests.

Two contrasting experimental groups were selected in exactly the same manner as Getzels and Jackson (i.e., a Hi-Low group (N = 17) and a Low-Hi group (N = 17)). A third experimental group consisted of the Hi-Hi's (N = 21). However, in this population, unselected as to IQ, the Hi-Low group was not significantly lower than the Low-Hi group on the DT total score.

The results showed that the Hi-Low group was significantly higher than the Low-Hi group on both achievement measures. The Hi-Hi group was significantly higher than the

---

27 Parween Hasan and H.J. Butcher, "Creativity and Intelligence: A Partial Replication with Scottish Children of Getzels' and Jackson's Study", British Journal of Psychology, Vol. 57, Nos. 1 and 2, 1966, p. 129-135.

Low-Hi group on the English achievement test. The authors do not state which statistical tests they used.

In this group therefore, the Getzels-Jackson phenomenon was not found. The authors account for the discrepancy by suggesting an IQ threshold, above which IQ would cease to be relevant to achievement and where DT tests would continue to be relevant.

Edwards and Tyler<sup>28</sup> tested the entire ninth grade of a non-selective public junior high school. Their IQ measure was the score on the School and College Achievement Test (SCAT). The achievement measures were Grade Point Average (GPA) and the average score on the Sequential Tests of Educational Progress (STEP). The DT measure was the total of two tests from the MTCT. The Hi-Low group consisted of 36 (22 girls and 14 boys) students who were in the top third on SCAT but not on the DT measure. The Low-Hi group consisted of 36 students (23 girls and 13 boys) who were in the top third on the DT measure but not on SCAT. A third group consisted of 26 students (15 boys and 11 girls) who were in the top third on both measures (the Hi-Hi group).

---

28 Meredith Payne Edwards and Leona E. Tyler, "Intelligence, Creativity and Achievement in a Nonselective Public Junior High School", Journal of Educational Psychology, Vol. 56, No. 2, 1965, p. 96-99.

An Analysis of Variance was performed to test for the effects of ability, sex, and ability-sex interaction between the Hi-Low and Low-Hi groups. The Hi-Low group was significantly superior on both criteria.

A second analysis of variance was used to compare the Hi-Low and Hi-Hi groups. There was no difference on the STEP, but the Hi-Hi group scored significantly lower on GPA than the Hi-Low group. In this case, it appears that high DT scores were a limiting factor in school achievement.

#### b. Multiple Correlation Studies

Torrance<sup>29</sup> tested 75 subjects, 25 in each grade from four through six, in a laboratory school. The IQ measure used was the Wechsler Intelligence Scale for Children (WISC). The DT measure was a total score from the MTCT battery. The achievement measures were the Gates Reading Test, and the Reading Skills, Study-Work Skills, English Skills and Arithmetic Skills of the Iowa Battery. Neither the method of sampling nor the mean IQ scores of the sample was reported. The WISC correlations with the five achievement measures ranged from .53 to .63. When the summated DT scores were added, the multiple correlations ranged from .57 to .68, with a median increase in the correlation coefficient of .07. The significance of the increases was not reported.

---

29 E. Paul Torrance, "Explorations in Creative Thinking in the Early School Years: VIII IQ and Creativity in School Achievement", Research Memorandum BER-59-11, Bureau of Educational Research, University of Minnesota, August 1969, p. 1-6.

Cline et al.<sup>30</sup> conducted a multiple correlational study of achievement in science, in which a battery of DT tests was added to an IQ test and the increase in correlation was examined. The sample consisted of 74 males and 40 females in a Utah urban high school. The sample was selected on the basis of having completed at least two courses in science beyond the introductory course and of having participated in an earlier study of divergent thinking.

The IQ measure used was the CTMM. The mean and SD on this test for both males and females were approximately 100 and 12. The DT measures used were 5 Structure-of-Intellect tests. The achievement measures were a science GPA and the STEP Science Test. Three other criteria were also used but since they were not strictly achievement measures they will not be discussed here. The data were analyzed separately for males and females. Only the results for males will be reported as the N for females was too low to expect a stable multiple correlation. When the DT battery was added to the IQ test the correlations with GPA rose from .61 to .70 and with the STEP Science Test they rose from .47 to .69.

---

<sup>30</sup> Victor B. Cline, James M. Richards Jr., and Walter E. Needham, "Creativity Tests and Achievement in High School Science", Journal of Applied Psychology, Vol. 47, No. 3, 1963, p. 184-189.

The author does not report the significance of these increases but concludes that the DT tests accounted for a substantial amount of criterion variance in addition to that accounted for by the IQ test.

Mullins<sup>31</sup> conducted a study in an attempt to discover how much is gained in predicting school performance by adding DT tests to a battery of standard aptitude tests, and conversely, how much is added to prediction when standard aptitude tests are added to a battery of DT tests.

The sample consisted of the entire student population (N = 225) of a high school in Texas. Although the mean IQ of the group was not reported it is possibly a selected group as the author reports that the average student who goes away to college maintains a B average<sup>32</sup>.

The standard aptitude measures were taken from the Project Talent battery. These were: Vocabulary, Visualization 2, Arithmetic Reasoning and Clerical Checking. The DT battery consisted of: Color-Form C, Word Meanings C, Brick Uses, the Omnibus Opinion Survey (a complexity scale), Imagery Clearness (a measure of Ego strength) and the Flanigan Creativity Tests. The criterion measure was GPA.

---

31 Cecil J. Mullins, "Current Studies of the Personnel Research Laboratory in Creativity" in Calvin W. Taylor, Editor, Widening Horizons in Creativity, New York, John Wiley and Sons, 1964, p. 320-326.

32 Idem, ibid., p. 324

The multiple R for the standard aptitude and DT batteries together was .74. For the aptitude battery alone it was .63 and for the DT battery alone it was .51. The difference between the full battery and the aptitude battery was not significant whereas the difference between the full battery and the DT battery was significant at the .001 level. That is, the DT battery when added to a standard aptitude battery did not significantly add to prediction. A limitation to this study is that the majority of the measures used do not seem to tap DT.

Guilford, Hoepfner and Peterson<sup>33</sup> studied the differential contribution of Structure of Intellect (SI) tests and standard aptitude tests to the prediction of success in mathematics in a grade 9 in a California school. There were four levels of mathematics taught: Basic Mathematics, Non-College Algebra, Regular Algebra, and Accelerated Algebra. The criterion measures were two specially constructed objective tests, a Basic Mathematics test for the first two levels and an Algebra test for the latter two groups. The N's ranged from 73 to 101.

---

<sup>33</sup> J.P. Guilford, Ralph Hoepfner and Hugh Peterson, "Predicting Achievement in Ninth Grade Mathematics from Measures of Intellectual Aptitude Factors", Educational and Psychological Measurement, Vol. 25, No. 3, 1965, p. 659-682.

There were 32 SI tests used: 4 DT tests, 8 tests in the Evaluation category, 11 in the Convergent-Production category, 8 in the Cognition category, and 1 in the Memory category. Twenty-five tests were administered in one testing session and seven in another. The first twenty-five tests were factor analyzed along with two standard aptitude tests (the Numerical Ability subtest of the DAT and the Language subtest of the CTMM) and 13 factors were obtained. Factor scores for each subject were obtained by summing the standard scores of the tests for each factor, but no test was used in more than one factor composite.

The standard aptitude measures were the Verbal Reasoning, Numerical Ability, Abstract Reasoning and Clerical Speed and Accuracy tests of the DAT; the CTMM Language MA and CTMM Non-Language MA; and three subtests of the Iowa Tests of Basic Skills, Reading Comprehension, Arithmetic Concepts and Arithmetic Problem Solving.

Multiple correlations were done on each of the four groups between the criterion scores and (1) the 13 SI factors, (2) the 7 SI tests, (3) the 13 SI factors plus 7 SI tests, (4) the 4 DAT tests, (5) the 2 CTMM tests, (6) the 3 Iowa tests, (7) the 9 standard aptitude tests (i.e., the DAT, the CTMM and the Iowa tests).

In the four groups, the 13 SI factors plus 7 SI tests and the 4 DAT tests broke just about even, with the SI tests

superior in predicting the results of the Regular Algebra group and the DAT superior in predicting the Basic Mathematics group.

Adding the 13 SI factor scores to the 4 DAT scores resulted in a significant improvement in the Regular and Accelerated Algebra groups only (from .29 to .55 and from .72 to .85 respectively).

The authors pointed out that examination of the beta weights showed that the few DT factors represented in the study had some indications of predictive contribution in all four courses<sup>34</sup>. They also concluded that the principle of improving prediction by bringing new factors into the regression equations is well supported<sup>35</sup>.

However, the obtained differences in predictive efficiency between the DAT tests and the SI factors was confounded by including some standard aptitude tests in the factor analyses. Further, the DAT measures take much less testing time and are more reliable than the SI measures.

None of the above studies attempted to cross-validate the regression equations.

---

34 Idem, ibid., p. 674.

35 Idem, ibid., p. 675.

### c. The Threshold Hypothesis

A third type of study was conducted by Cicirelli<sup>36</sup>. The research was designed to test the hypothesis that IQ has an effect on academic success up to a certain point but that after this threshold is passed DT abilities would begin to have an effect. He tested 641 sixth grade pupils in a suburban Detroit school system.

The DT measures were the MTCT. The standard aptitude measure was the CTMM short form. The three achievement measures were the California Arithmetic Test, the California Language Test and the Gates Basic Reading Tests. The population was categorized into eight levels of intellectual ability and three levels of DT ability. The IQ scores of the population ranged from 70-162.

The results showed linearity and additivity rather than thresholds and interaction. Correlations between the composite DT score and the 3 criterion measures with IQ partialled out ranged from .07 to .30.

### 3. The Problem and Hypothesis

In general, the relationship of divergent thinking to academic success is still obscure.

---

<sup>36</sup> Victor G. Cicirelli, "Form of the Relationship between Creativity, IQ and Academic Achievement", Journal of Educational Psychology, Vol. 56, No. 6, 1965, p. 303-308.

The evidence from studies using the extreme group design although contradictory, gives some indication that DT tests may make some contribution to the prediction of academic success. The multiple correlation studies are a superior design as they permit the assessing of the precise amount of criterion variance accounted for by DT tests independent of standard aptitude tests. However, the nature of the technique imposes two basic demands on the investigator. The first is a sufficiently large sample size to insure the stability of the multiple correlation. The second is a cross-validation sample. These conditions are necessary because the multiple correlation technique capitalizes upon chance relationships in the sample data and hence overestimates the true correlation<sup>37</sup>. None of the studies reviewed employed a cross-validation sample and some employed small sample sizes, i.e., N's of under 100.

A multiple correlation study could help determine whether DT tests account for any sizeable incremental validity over standard aptitude tests in the prediction of academic success. Certain criteria would have to be met: the standard aptitude battery should contain those tests shown to have the greatest predictive validity -- a general intelligence test plus verbal reasoning and numerical ability

---

<sup>37</sup> Paul A. Hertzberg, "The Parameters of Cross-Validation", Psychometrika, Vol. 34, No. 2, Part 2, June 1969, p. 1.

aptitude tests; the criterion measures should be univocal in that the entire sample is measured on the same test; the sample should be sufficiently heterogeneous so that the correlations will not be attenuated through restriction of range; the DT tests should be chosen according to a rationale other than simple availability; and a cross-validation sample should be employed.

The next chapter presents the experimental design for such a study. It is intended to test the null hypothesis:

There is no significant increase in the cross-validated prediction of academic success by adding some divergent thinking tests to some standard aptitude tests in a multiple regression equation.

## CHAPTER II

### EXPERIMENTAL DESIGN

This chapter outlines the sampling procedures, the measures used, the data gathering procedures, the reliability measures, the method of statistical analysis and the subsequent statistical manipulations employed in this study.

#### 1. The Sample

Virtually all freshmen at Sir George Williams University, Montreal take the Introductory Psychology Course<sup>1</sup>. At registration, they are alternately assigned to one of three sections, according to their order of registration. One of these sections in the 1967-1968 academic year was chosen as the experimental group (hereafter called the Validation group). The same section was chosen in the 1968-1969 academic year for cross-validation (hereafter called the Cross-validation group). To obtain a univocal criterion measure only those freshmen who had graduated from a Protestant High School in the province of Quebec were selected.

The 120 subjects in the Validation sample came from 28 different High Schools. There were 58 males and 62 females with a mean age of 18.6 years. One hundred eight subjects listed English as their first language.

---

1 T. Swift, Personal communication.

The 109 subjects in the Cross-validation sample came from 27 of the same 28 High Schools as the Validation sample. There were 49 males and 60 females with a mean age of 17.3 years. Ninety-seven subjects listed English as their first language.

## 2. The Measuring Instruments

### a. Criterion Measures

Since all freshmen at Sir George Williams University take two courses in common, Introductory Psychology and English Composition, it was originally intended to use these course grades as the criterion measures. However, subsequent examination showed that the measuring units were too coarse and too skewed to be useful for prediction. That is, letter rather than numerical grades were given and in one course, English Composition, 75% of the students received a grade of C.

Consequently, the students' 1967 High School Matriculation results were used as the criterion. All students took the same five examinations: English Grammar and Composition (GRA), English Literature (LIT), French Grammar and Composition (WRI), Oral French (ORA) and History (HIS). Since the sample was already attending University, it obviously includes only those who succeeded in the examinations and who possessed the motivation and means to go to University. Nevertheless, as will be shown in the next

chapter, the range of scores on all variables was not unduly restricted. The Cross-validation sample wrote the 1968 edition of the same five examinations.

b. Predictor Measures

The standard aptitude measures were chosen to include that type of test which, in the review of the literature, were found to be most effective in predicting academic success. These were the VR, NR and AR subtests of the DAT (form L) and the Gamma form C of the Otis SA Tests of Mental Ability. Although the AR subtest of the DAT was not found to be a particularly useful predictor of academic success, it was included because of its figural content. Thereby, semantic, symbolic and figural content areas were included. This covers the same content areas as the DT tests.

The DT measures were chosen on the basis of several criteria: all content categories should be represented; all product categories should be represented; the test materials should elicit responses that are scorable without undue reliance on personal judgement and hence will yield high interscorer reliability; the test should have at least two separately timed parts to allow for calculating split-half reliability; a practical restraint was the  $1\frac{1}{4}$  hr. time limit on testing which was the length of the class time available.

The DT tests chosen, on the basis of the above criteria, were:

- (1) Word Beginnings and Endings (DSU) - 2 parts;  
2 minutes per part. Sample question: Write as many words as you can beginning with S and ending with T.
- (2) Alternate Uses (DMC) - 3 parts; 4 minutes per part. Sample question: List as many uses as you can for a door handle.
- (3) Match Problems (DFT) - 2 parts; 7 minutes per part. Sample question: Given a figure consisting of a number of squares made with matchsticks, take away a certain number of matchsticks leaving a certain number of squares.
- (4) Associational Fluency (DMR) - 2 parts; 4 minutes per part. Sample question: Write as many words as you can similar in meaning to the word hard.
- (5) Expressional Fluency (DMS) - 4 parts; 2 minutes per part. Sample question: Write as many sentences as you can, using the following four letters as the first letter of each word:  
W \_\_\_\_\_ f \_\_\_\_\_ r \_\_\_\_\_ d \_\_\_\_\_.

The tests will hereafter be referred to by their trigram code. The first letter, which is always D represents the operation of Divergent Production. The second letter

represents the content category, M for semantic content, F for figural content and S for symbolic content. The last letter represents the products category, U for Units, C for Classes, T for Transformations, R for Relations and S for Systems. Guilford's system is elaborated in a recent publication<sup>2</sup>.

All content categories are represented except the Behavioral which was not available at the time of testing. All product categories are represented except Implications which was not included due to the time constraint.

### 3. The Data Gathering Procedures

For the Validation group, the criterion measures were written by the subjects in May 1967. The standard aptitude measures were written in July 1967. Scores were obtained from the University records. The DT tests were written in October 1967 during a 1½ hour lecture period. The subjects were seated in every second seat in a well lighted lecture auditorium. There were 12 proctors. A similar schedule was observed in 1968 in regard to the Cross-validation group.

---

<sup>2</sup> J.P. Guilford, The Nature of Human Intelligence, New York, McGraw-Hill, 1967, p. 60-66.

Despite the fact that the criterion measures were gathered prior to the predictor measures the term prediction rather than postdiction is being used according to the rationale set down by Nunnally.

The term prediction will be used in a general (and ungrammatical) sense to refer to functional relations between an instrument and events occurring before, during, and after an instrument is applied. Thus a test administered to adults could be used to make "predictions" about events occurring in their childhood... Others have referred to predictive validity at ... three points in time, respectively, as "postdiction", "concurrent validity", and "prediction". Using different terms however suggests that the logic and procedures of validation are different which is not true. In each case a predictor measure is related to a criterion measure, and after the data are available it does not matter when they were obtained.<sup>3</sup>

#### 4. Reliability

The general level of reliability of the aptitude and intelligence tests used is well established. The Manual for the Otis Gamma test reports an odd-even reliability of .85 on a Grade 12 population in New York State where the N was 84<sup>4</sup>. A split-half reliability estimate on a Grade 12 population for form L of the DAT tests used is reported in Table I.

---

<sup>3</sup> Jum C. Nunnally, Psychometric Theory, New York, McGraw-Hill, 1967, p. 76-77.

<sup>4</sup> Arthur S. Otis, Otis Quick Scoring Mental Ability Tests: Manual of Directions for Gamma Test, Forms C and D, New York, Harcourt, Brace and World, 1939, p. 6.

TABLE I.- Split-Half Reliability Coefficients<sup>a</sup> of the  
DAT Subtests on a Grade 12 Population

Test	Boys N : 168	Girls N : 179
VR	.94	.93
NA	.93	.91
AR	.92	.94

a George K. Bennett, Harold G. Seashore and Alexander G. Wesman, Manual for the Differential Aptitude Tests Forms L and M, New York, The Psychological Corporation, 1966, p. 6-2.

The reliability of the DT tests was estimated in two ways. Split-half reliability was calculated by intercorrelating the scores for the separately timed parts of each test and extending this with the Spearman-Brown prophecy formula. Test-retest reliability was estimated by retesting 23 paid volunteers 6 weeks after the initial testing session and correlating their scores. (All correlations reported in this study are Pearson product-moment correlations.) The estimated reliabilities are reported in Table II, the intercorrelations among the part scores in Table III and the standard deviations of the parts of the DT tests in Table IV. From Table IV it can be seen that the SD's of the part scores are of sufficiently uniform size to permit the calculation of the Spearman-Brown formula.

Interscorer reliability of the DT tests was established in the following manner: when the test booklets had been placed in alphabetical order, every third booklet was extracted. The author and a second scorer, who had previously jointly scored 10 booklets from a different group, independently scored the 40 selected booklets. The two sets of scores were then

TABLE II.- Reliabilities of the DT Tests

Test	Split-Half N : 120	Test-Retest N : 23
DSU	.72	.51
DMC	.73	.70
DFT	.55	.58
DMR	.37	.33
DMS	.45	.35

TABLE III.- Intercorrelations Among Parts  
of the DT Tests N : 120

Test	Parts	r
DSU	1 and 2	.55
DMC	1 and 2	.44
	1 and 3	.48
	2 and 3	.50
DFT	1 and 2	.38
DMR	1 and 2	.37
DMS	1 and 2	.15
	1 and 3	.10
	1 and 4	.07
	2 and 3	.28
	2 and 4	.23
	3 and 4	.17

TABLE IV.- Means and Standard Deviations of the  
Parts of the DT Tests N : 120

Test	Mean	SD
DSU - Part 1	11.3	3.8
Part 2	7.6	3.0
DMC - Part 1	6.9	2.7
Part 2	7.3	2.1
Part 3	6.9	2.2
DFT - Part 1	5.4	2.0
Part 2	4.3	1.9
DMR - Part 1	6.1	2.2
Part 2	5.2	2.3
DMS - Part 1	1.4	0.9
Part 2	1.4	1.0
Part 3	1.5	1.0
Part 4	1.4	1.1

correlated. Results are presented in Table V. The coefficients are uniformly high which is to be expected as one of the criteria for choosing the tests was minimal scorer judgement.

Since the criterion measures were Matriculation results it was not possible to estimate reliability in any of the conventional ways. In order to give some indication of the consistency of scores a correlation was computed between the mid-year and matriculation scores for each subject. These are reported in Table VI. It is realized that the content of the two sets of examinations was not identical thereby probably attenuating the obtained correlations.

## 5. Statistical Analysis

Stepwise multiple correlations with forced entry of variables<sup>5</sup> were computed using the standard aptitude tests as predictor variables against each criterion measure in turn. The same procedure was then employed using all 9 predictor variables. The difference in multiple R's between the standard aptitude battery alone and the standard aptitude battery plus the DT tests is a measure of the amount of

---

<sup>5</sup> W.J. Dixon, Editor, BMD Biomedical Computer Programs, Berkeley, University of California Press, 1967, p. 233.

TABLE V.- Interscorer Reliability  
on the DT Tests N : 40

Test	r
DSU	.99
DMC	.91
DFT	.92
DMR	.84
DMS	.93

TABLE VI.- Correlations Between Mid-Year  
and Matriculation Scores N : 120

Test	r
GRA	.55
LIT	.62
WRI	.68
ORA	.66
HIS	.48

criterion variance associated with the DT tests independent of the standard aptitude tests (incremental validity). The null hypothesis of no differences between the cross-validated Multiple R's was tested by the formula<sup>6</sup>:

$$F = \frac{(R_1^2 - R_2^2) (N - m_1 - 1)}{(1 - R_1^2) (m_1 - m_2)}$$

where  $R_1$  = multiple R with larger number of independent variables

$R_2$  = multiple R with one or more variables omitted

$m_1$  = larger number of independent variables

$m_2$  = smaller number of independent variables

A stepwise Multiple R was employed for two reasons. First, a computer program was available at no cost and secondly, for future research a subset of predictors might be more practical than administering all 9 predictor tests.

However, since the above mentioned F test requires that the multiple R with the larger number of variables include the same variables as the multiple R with the smaller number of variables, forced entry of variables was used. This yields the same magnitude multiple R as the full model multiple regression.

---

<sup>6</sup> J.P. Guilford, Fundamental Statistics in Psychology and Education, 3rd Edition, New York, McGraw-Hill, 1956, p. 400.

## 6. Subsequent Analyses

Since the prediction equations did not cross-validate, subsequent analyses were performed. To test whether the Validation and Cross-Validation samples come from different populations in terms of their scores on the tests, a Discriminant Analysis<sup>7</sup> was performed using all 14 tests (i.e., both predictor and criterion measures).

As the groups proved to be significantly different, two new samples were constructed by randomly splitting each group in half and placing the first half of the Validation group with the first half of the Cross-Validation group and following a similar procedure for the second half of each group. A second Discriminant Analysis was performed and the two new groups were not found to be significantly different. The two new groups thereby formed are called the Calibration and Verification groups. A double cross-validation design was then performed. That is, equations derived from the Calibration group were cross-validated against the Verification group and vice-versa.

---

7 Dixon, op. cit., p. 185

## 7. Statistical Assumptions

The correlation technique assumes linearity of regression and homoscedasticity<sup>8</sup>.

The assumption of linearity of regression was tested by comparing the Pearson r's between each predictor and criterion variable with the corresponding correlation ratio, using the formula<sup>9</sup>:

$$F = \frac{(\eta^2 - r^2)(N - k)}{(1 - \eta^2)(k - 2)}$$

where k = the number of columns

The results of this analysis are presented in Appendix 1. One F test was found to be significant at the .05 level. Since in 45 tests this could have occurred by chance, all regressions are considered to be linear.

Regarding homoscedasticity, Guilford states:

When columns (and rows) are relatively homoscedastic, we may compute a Pearson r. This condition will prevail generally when the two distributions are fairly symmetrical within themselves, thus we need not go so far as to compute standard deviations of columns and rows in order to find out. It is when distributions are markedly skewed that significant departures from homoscedasticity occur.<sup>10</sup>

---

8 Guilford, op. cit., p. 149.

9 Idem, ibid., p. 294.

10 Idem, ibid., p. 150.

Consequently, the distributions of all tests were plotted graphically. The most deviant appearing distribution was tested by the Chi-Square test for goodness of fit<sup>11</sup>. If this proved to be significant the scores were normalized to a T-score distribution according to the method described by Ghiselli<sup>12</sup>. Then, the next most deviant appearing distribution was tested until a non-significant Chi-Square was found. The test distributions that required normalization were NA and HIS. The AR test was also normalized due to a subsequently corrected error in calculating the Chi-Square.

---

11 Idem, *ibid.*, p. 240-242.

12 Edwin E. Ghiselli, *Theory of Psychological Measurement*, New York, McGraw-Hill, 1964, p. 84-88.

## CHAPTER III

### RESULTS AND DISCUSSION

#### 1. Description of the Data

The mean, standard deviation, maximum score, minimum score and range of each variable are presented in Tables VII and VIII. The Validation sample, compared to the DAT normative population of Grade 12 boys (Form L) scored from the 35th to 97th centiles on VR, from the 25th to 99th centiles on NA and from the 10th to the 97th centiles on AR<sup>1</sup>. It can be seen that the scores are not unduly restricted in range. Similar comparisons among the other tests are not available. The Cross-Validation sample shows a similar range of scores.

#### 2. Validation and Cross-Validation Results

Table IX presents the intercorrelations among all the tests for the Validation and Cross-Validation groups. The intercorrelations among the predictor tests are satisfactorily low as none are above .50. Similarly, the highest correlation among the criterion tests is .54 between GRA and LIT, preventing the summation of the criteria into a GPA.

---

<sup>1</sup> George K. Bennett, Harold G. Seashore and Alexander G. Wesman, Manual for the Differential Aptitude Tests, 4th Edition, New York, The Psychological Corporation, 1966, p. 3-12.

TABLE VII.- Data Description of the  
Validation Sample N : 120

Variable	Mean	S.D.	Maximum Score	Minimum Score	Range
OTIS	57.7	6.7	73	42	32
VR	39.2	4.9	48	26	23
NA	32.9	4.0	40	19	22
AR	39.7	4.5	47	22	26
DSU	19.0	6.0	33	2	32
DMC	21.1	5.6	38	9	30
DFT	9.7	3.2	19	2	18
DMR	11.3	3.7	23	3	21
DMS	5.7	2.5	12	1	12
GRA	68.0	11.1	92	34	59
LIT	68.1	10.0	87	37	51
WRI	68.3	11.3	94	35	60
ORA	68.1	10.5	99	35	65
HIS	70.2	10.3	94	32	63

TABLE VIII.- Data Description of the  
Cross-Validation Sample N : 109

Variable	Mean	S.D.	Maximum Score	Minimum Score	Range
OTIS	54.2	7.1	73	40	34
VR	40.6	4.6	49	24	26
NA	33.4	3.7	40	21	20
AR	39.7	4.9	49	27	23
DSU	19.9	5.5	33	6	28
DMC	20.7	5.7	36	5	32
DFT	9.4	3.6	19	3	17
DMR	15.1	5.1	36	5	32
DMS	7.1	3.4	19	1	19
GRA	69.4	9.0	92	43	50
LIT	69.1	8.1	90	43	48
WRI	66.5	9.4	91	40	52
ORA	66.3	11.6	98	28	71
HIS	67.1	9.9	87	40	48

TABLE IX.- Intercorrelations Among All the Tests on the Validation and Cross-Validation Groups

Validation Group	Cross-Validation Group													
	1 OTIS	2 VR	3 NA	4 AR	5 DSU	6 DMC	7 DFT	8 DMR	9 DMS	10 GRA	11 LIT	12 WRI	13 ORA	14 HIS
1. OTIS		.45	.50	.37	.21	-.07	.07	.32	.16	.07	.08	.14	.13	.04
2. VR	.33		.23	.33	.31	.12	-.08	.34	.27	.38	.18	.28	.23	.02
3. NA	.19	-.16		.31	.10	-.10	.08	.16	.15	-.12	-.05	.05	-.02	.10
4. AR	.48	.18	.20		.20	-.03	.18	.09	.03	.08	.11	.22	.13	-.05
5. DSU	.03	.18	-.18	-.02		.18	.12	.32	.35	.08	.13	.17	.01	-.10
6. DMC	.18	.11	-.03	.04	-.01		-.07	.22	.17	.04	.07	-.18	-.04	-.05
7. DFT	.43	.03	.21	.46	-.02	.01		-.05	.01	-.11	-.08	.03	.00	-.09
8. DMR	.02	.21	-.17	-.18	.43	.14	-.04		.38	.25	.26	.21	.06	.06
9. DMS	.05	.18	-.19	.12	.39	.16	-.05	.34		.10	.15	.15	-.03	-.03
10. GRA	.17	.40	-.35	.04	.37	.06	-.18	.36	.32		.46	.49	.37	.12
11. LIT	.15	.43	-.24	.09	.12	.07	-.16	.29	.26	.54		.28	.27	.11
12. WRI	-.10	.10	-.16	-.22	.39	-.04	-.28	.25	.22	.46	.22		.51	.13
13. ORA	-.15	-.01	-.11	-.18	.35	.02	-.20	.24	.18	.28	.09	.59		-.07
14. HIS	-.09	.06	-.14	-.22	.20	-.03	-.14	.23	.14	.25	.11	.25	.10	

Table X shows the multiple stepwise correlations with forced entry of variables between the two sets of predictors and all five criteria.

The significance of all Multiple R's was established by reference to Guilford's table<sup>2</sup>. The significance of the F tests was read from a table of the distribution of  $F^3$ . In this case, since the appropriate table did not give the values for the precise number of degrees of freedom in the sample, the closest approximation was used. The df for the Validation sample are 5,110 and for the Cross-Validation sample are 5,99. In each case the closest approximation is df 5,100.

As seen in Table X, the Multiple R's between the aptitude subset and the five predictors were significant for the GRA and LIT tests only. The Multiple R's between the full set of predictors and the criteria were significant at the .01 level for all except the HIS test. The significance of the increase for all criteria except HIS was also significant at the .01 level. Hence it would appear that the DT tests added significantly to prediction in four of the five criteria. However, when the regression equations

---

2 J.P. Guilford, Fundamental Statistics in Psychology and Education, 3rd Edition, New York, McGraw-Hill, 1956, p. 539.

3 Idem, ibid., p. 542.

TABLE X.- Multiple R's and Cross-Validation R's

	Multiple R N : 120			Cross-Validation R N : 109		
	Aptitude Only	Aptitude and D.T.	F	Aptitude Only	Aptitude and D.T.	F*
GRA	.51**	.63**	6.24**	.37**	.33	-0.78
LIT	.47**	.56**	3.71**	.20	.27	0.88
WRI	.28	.50**	6.29**	-.15	.07	-0.44
ORA	.21	.42**	4.42**	-.03	.01	-0.02
HIS	.26	.34	1.65	-.05	.01	-0.06

\*\* p ≤ .01

.99 F 5,100 = 3.20

\* p ≤ .05

.95 F 5,100 = 2.30

were applied to the Cross-Validation sample (Table X, fifth and sixth columns) only one relationship remained significant, that between the Aptitude subset and the GRA criterion test. In fact, the cross-validation R of the smaller set of predictors (the Aptitude subset) was higher than that of the full set of predictors, .37 versus .33. Since there were no significant differences on Cross-Validation, the null hypothesis was accepted. The question of why the predictors did not cross-validate remains to be considered. Four possibilities will be discussed.

Possibility one: In many cross-validation procedures, a sample is randomly split into validation and hold out groups. This assures the equivalency of the two groups on the measures used. However, in this study, the Cross-Validation group was comprised of students entering the University the year after the Validation group, a more realistic prediction situation. Hence, the Cross-Validation group is only assumed to be similar to the Validation group. The subjects may, in fact, differ widely in their performance on both predictor and criterion variables.

Possibility two: The criterion tests written by the Cross-Validation group are only theoretically equivalent to those written by the Validation group. That is, there may be important differences in the construction of Academic Achievement tests between one year and the next.

Possibility three: Either predictor or criterion measures or both may not be sufficiently reliable to allow for cross-validation.

Possibility four: The technique itself may be responsible.

These possibilities will now be considered.

In order to test for possibility one, differences between the two groups in performance on the tests, a Discriminant analysis using all 14 measures was performed. The Mahalanobis D Square was 2.563 and the associated F test with 14 and 214 degrees of freedom was 9.86. Since .99 F 14,200 is 2.17, the obtained F is highly significant.

In order to compensate for this difference and also for possibility two, that the criterion tests are not equivalent, the two groups were randomly divided and reconstituted as described in Chapter II. The two new groups thus created are called the Calibration and Verification groups. A Discriminant analysis showed no significant differences between these two groups. The Mahalanobis D Square is .248 and the associated F is .955.

A Multiple Correlation analysis similar to that done on the first two groups was carried out. However, in this case a double cross-validation designed was used. That is, the prediction equation generated on the Calibration group was cross-validated on the Verification group and vice versa.

The zero order correlations are presented in Table XI and the Multiple R's in Tables XII and XIII. Table XI is quite similar to Table IX, and the same comments apply.

When Tables XII and XIII are compared to Table X, virtually the same pattern is found. That is, the difference between the full set of predictors and the aptitude subset was significant on some of the criteria but these differences did not hold up under cross-validation. Again as in Table X only prediction of the GRA criterion held up under cross-validation.

The cross-validated prediction of the LIT test from the aptitude subset of .39 was significant when the Verification group was predicted from the Calibration group but was reduced to a non-significant .26 when the Calibration group was predicted from the Verification group. Therefore, cross-validation cannot be said to have been achieved in this case.

The failure to cross-validate cannot be attributed to differences in mean scores between the Validation and Cross-Validation groups.

Possibility three, that the failure to cross-validate is due to the unreliability of the instruments cannot be tested, it can only be discussed. Guilford suggests that

If a composite from a battery is to be used ... it is likely that there is not too much to be gained by achieving reliabilities for single tests higher than .60 ...<sup>4</sup>

---

4 Idem, ibid., p. 472.

TABLE XI.- Intercorrelations Among All the Tests  
on the Calibration and Verification Groups

Calibration Group	Verification Group													
	1 OTIS	2 VR	3 NA	4 AR	5 DSU	6 DMC	7 DFT	8 DMR	9 DMS	10 GRA	11 LIT	12 WRI	13 ORA	14 HIS
1. OTIS		.45	.30	.40	.11	.11	.35	.07	.00	.07	.14	.01	-.01	-.09
2. VR	.27		.13	.22	.18	.07	.10	.23	.16	.30	.39	.09	.00	.03
3. NA	.36	-.03		.17	-.16	.00	.08	.03	-.09	-.37	-.23	-.17	-.08	.01
4. AR	.41	.24	.30		.03	-.03	.38	-.16	-.20	.03	.22	-.03	.02	-.08
5. DSU	.09	.30	.05	.13		.04	.12	.28	.32	.21	.07	.11	.19	.04
6. DMC	.01	.12	-.11	.01	.10		.04	.16	.27	.10	-.02	-.15	-.08	-.06
7. DFT	.16	-.11	.22	.25	-.02	-.10		-.08	-.18	-.10	.01	-.20	-.04	-.13
8. DMR	.08	.34	.05	.07	.44	.13	-.04		.39	.23	.23	.16	.10	.14
9. DMS	.09	.30	.07	.26	.44	.02	.10	.44		.19	.14	.09	-.02	-.07
10. GRA	.15	.45	-.11	.07	.28	-.02	-.16	.35	.27		.47	.35	.26	.20
11. LIT	.06	.29	-.08	-.01	.19	.13	-.25	.30	.22	.57		.17	.11	.13
12. WRI	.07	.23	.04	.00	.40	-.01	-.09	.17	.25	.50	.25		.45	.14
13. ORA	.06	.18	-.01	-.08	.17	.09	-.15	.08	.12	.36	.22	.48		-.01
14. HIS	.02	.09	-.03	-.13	.10	-.05	-.12	.12	.16	.24	.07	.20	.03	

RESULTS AND DISCUSSION

TABLE XII.- Multiple R's and Cross-Validation R's  
Verification Group Predicted From  
the Calibration Group

	Multiple R N : 115			Cross-Validation N : 114		
	Aptitude Only	Aptitude and D.T.	F	Aptitude Only	Aptitude and D.T.	F*
GRA	.47**	.55**	3.07*	.40**	.45**	1.38
LIT	.31*	.44**	3.17*	.39**	.27	-2.22
WRI	.25	.45**	4.61**	.04	.15	0.55
ORA	.22	.30	1.20	-.03	-.02	-0.01
HIS	.19	.29	1.38	.06	.05	-0.03

\*\*  $p \leq .01$

.99  $F_{5,100} = 3.20$

\*  $p \leq .05$

.95  $F_{5,100} = 2.30$

TABLE XIII.- Multiple R's and Cross-Validation R's  
Calibration Group Predicted From the  
Verification Group

	Multiple R N : 114			Cross-Validation R N : 115		
	Aptitude Only	Aptitude and D.T.	F	Aptitude Only	Aptitude and D.T.	F
GRA	.51**	.56**	2.03	.40**	.46**	1.72
LIT	.51**	.57**	2.49*	.26	.29	0.47
WRI	.21	.37	2.80*	.11	.21	0.88
ORA	.09	.26	1.66	-.02	.10	0.25
HIS	.14	.27	1.49	.11	.06	-0.22

\*\*  $p \leq .01$

.99  $F_{5,100} = 3.20$

\*  $p \leq .05$

.95  $F_{5,100} = 2.30$

There is little doubt about the reliabilities of the Otis and the DAT subtests. The reliabilities of the DT tests are more suspect. The split-half reliabilities, which range from .45 to .73 are of borderline value, but two of the test-retest reliabilities are in the .30's and none is higher than .70. Although the test-retest reliabilities are based on an N of only 23, they are consistent with the findings of Stoker and Kropp<sup>5</sup>. Those authors retested approximately 70 Grade 12 students ten months after the initial testing on a series of tests from the Educational Testing Service's Kit of Cognitive Factors, many of which are similar to Guilford's Structure of Intellect tests. They found a median test-retest reliability of .56 on 38 tests. Three of their tests measured three of the DT factors used in this study. The reliability of DFT was .73, of DMR .23, and of DMS .40.

The reliabilities of the criterion tests range from .48 to .68. As has already been discussed in Chapter II, these are only approximations to a true reliability estimate. Nevertheless, essay type questions generally are not very reliable so the criterion tests probably bear some responsibility for the failure to obtain significant results upon cross-validation.

---

5 H.W. Stoker and R.P. Kropp, "Note on the Kit of Reference Tests", Educational and Psychological Measurement, Vol. 27, No. 4, 1967, p. 1171-1172.

The fourth possibility is that, as has already been mentioned, the technique capitalizes on chance and gives a biased estimate of the population multiple correlation which is then deflated on cross-validation. According to Hertzberg<sup>6</sup> this bias tends to decrease with increasing sample size and to increase with an increase in the number of predictors. Consequently, an increase in sample size would be recommended for any future study with nine predictor variables.

### 3. Summary and Discussion

In all of the analyses undertaken, a significant cross-validated prediction was possible only with the GRA criterion. In the case of this criterion, the null hypothesis of no differences in predictive power between the aptitude subset of predictors and the full set of predictors was accepted.

The failure of the majority of prediction equations to cross-validate was discussed. A larger sample and more reliable measures may prevent this from happening in any future study.

---

<sup>6</sup> Paul A. Hertzberg, "The Parameters of Cross-Validation", Psychometrika, Vol. 34, No. 2, Part 2, June 1969, p. 65.

The question then of whether the DT tests used have any external validity or practical usefulness in academic prediction is still an unsettled one. What emerges from this study is that in the one criterion where a significant cross-validated prediction was possible, the DT tests did not add significant incremental validity to a battery of standard aptitude tests.

This study also illustrates the importance of cross-validation in prediction studies, a rule more honored in the breach than in the observance in the DT literature.

A future multiple correlation study bent on examining the practical usefulness of Guilford's DT tests in academic prediction should have several features: a well chosen criterion test with high reliability; a larger and more representative DT battery than was used in the present study and one in which the length of some tests is increased to ensure better reliability, and a large sample in both validation and cross-validation groups.

## BIBLIOGRAPHY

Cicirelli, Victor G., "Form of the Relationship Between Creativity I.Q. and Academic Achievement", Journal of Educational Psychology, Vol. 56, No. 6, 1965, p. 303-308.

This article investigated the "threshold hypothesis". It was found that the form of the relationship between IQ, Divergent Thinking and Academic Achievement was linear and additive.

Cline, Victor B., James M. Richards Jr., and Walter E. Needham, "Creativity Tests and Achievement in High School Science", Journal of Applied Psychology, Vol. 47, No. 3, 1963, p. 184-189.

This is a Multiple Regression study which found that Divergent Thinking tests accounted for a substantial amount of criterion variance in addition to that accounted for by an IQ test.

Edwards, Meredith Payne and Leona E. Tyler, "Intelligence, Creativity and Achievement in a Nonselective Public Junior High School", Journal of Educational Psychology, Vol. 56, No. 2, 1965, p. 96-99.

An extreme group study which found the High IQ, Low Divergent Thinking group to be significantly superior to all other groups, including a High IQ, High Divergent Thinking group.

Flescher, Irwin, "Anxiety and Achievement of Intellectually Gifted and Creatively Gifted Children", Journal of Psychology, Vol. 56, Second Half, 1963, p. 251-268.

An extreme group study which found that IQ was related to academic success but that Divergent Thinking was not.

Getzels, Jacob W., and Philip W. Jackson, Creativity and Intelligence, New York, John Wiley and Sons, 1962, xvii-293 p.

A landmark publication which generated much research into the nature of the relationship among IQ, Divergent Thinking and academic achievement. Informative clinical and anecdotal results are marred by weaknesses in design and statistical inference.

Guilford, J.P., Ralph Hoepfner and Hugh Petersen, "Predicting Achievement in Ninth-Grade Mathematics from Measures of Intellectual Aptitude Factors", Educational and Psychological Measurement, Vol. 25, No. 3, 1965, p. 659-682.

An extensive study of the relationship between Structure of Intellect tests and mathematical achievement. The Divergent Thinking tests had significant beta weights in many of the prediction equations.

Guilford, J.P., The Nature of Human Intelligence, New York, McGraw-Hill, 1967, vii-538 p.

A major publication in which the author fully discusses his theory, his tests and his research results.

Hasan, Parween, and H.J. Butcher, "Creativity and Intelligence: A Partial Replication with Scottish Children of Getzels' and Jackson's Study", British Journal of Psychology, Vol. 57, Nos 1 and 2, 1966, p. 129-135.

An extreme group study which found that IQ was related to academic achievement but that Divergent Thinking was not.

McNemar, Quinn, "Lost: Our Intelligence? Why?", American Psychologist, Vol. 19, No. 12, 1964, p. 871-882.

A vivid critique of theories and tests in the realm of intellectual abilities. The comparative validities of different genera of tests are evaluated.

Mullins, Cecil J., "Current Studies of the Personnel Research Laboratory in Creativity" in Calvin W. Taylor, Editor, Widening Horizons in Creativity, New York, John Wiley and Sons, 1964, xix-466 p.

A Multiple Correlation study which found that Divergent Thinking did not add significantly to academic prediction. However, it is dubious whether the measures used tap Divergent Thinking as defined by any of the major theoretical positions.

Torrance, E. Paul, "Explorations in Creative Thinking in the Early School Years: VIII IQ and Creativity in School Achievement", Research Memorandum BER-59-11, Bureau of Educational Research, University of Minnesota, August 1969, 6 p.

A Multiple Correlation study in which a Divergent Thinking battery, when added to an IQ test, produced a small increase in prediction.

-----, Guiding Creative Talent, Englewood Cliffs, Prentice-Hall, 1962, xi-278 p.

A major source of the author's theory, early test rationale and research. A series of extreme group studies are reported, the majority of which found Divergent Thinking to be related to academic achievement.

Torrance, E. Paul, "Examples and Rationales of Test Tasks for Assessing Creative Abilities", Journal of Creative Behavior, Vol. 2, No. 3, 1968, p. 165-178.

This article reports on the continued development of the Minnesota Tests of Creative Thinking.

Yamamoto, Kaoru, "Role of Creative Thinking and Intelligence in High School Achievement", Psychological Reports, Vol. 14, 1964, p. 783-789.

An extreme group study which found Divergent Thinking scores to be related to academic achievement.

-----, "A Further Analysis of the Role of Creative Thinking in High School Achievement", Journal of Psychology, Vol. 58, First Half, 1964, p. 277-283.

A re-analysis of the previous data which showed that when IQ was controlled, subjects high on Divergent Thinking were significantly better academic achievers than subjects low on Divergent Thinking.

APPENDIX 1

TESTS FOR LINEARITY OF REGRESSION

TABLE XIV.- Regression of the GRA Criterion  
on the Nine Predictors

Test	r	eta yx	No. of Columns	df	F
OTIS	.17	.30	12	10,108	0.73
VR	.40	.45	12	10,108	0.58
NA	-.35	.41	12	10,108	0.59
AR	.04	.29	12	10,108	0.97
DSU	.37	.47	12	10,108	1.16
DMC	.06	.39	12	10,108	1.89
DFT	-.18	.29	12	10,108	0.61
DMR	.36	.41	12	10,108	0.50
DMS	.32	.40	12	10,108	0.74

TABLE XV.- Regression of the LIT Criterion  
on the Nine Predictors

Test	r	eta yx	No. of Columns	df	F
OTIS	.15	.29	12	10,108	0.73
VR	.43	.48	12	10,108	0.64
NA	-.24	.38	12	10,108	1.10
AR	.09	.35	12	10,108	1.41
DSU	.12	.25	12	10,108	0.55
DMC	.07	.32	12	10,108	1.17
DFT	-.16	.31	12	10,108	0.84
DMR	.29	.36	12	10,108	0.56
DMS	.26	.33	12	10,108	0.50

TABLE XVI.- Regression of the WRI Criterion  
on the Nine Predictors

Test	r	eta yx	No. of Columns	df	F
OTIS	-.10	.25	12	10,108	0.60
VR	.10	.28	12	10,108	0.80
NA	-.16	.31	12	10,108	0.84
AR	-.22	.29	12	10,108	0.42
DSU	.39	.50	12	10,108	1.41
DMC	-.04	.37	12	10,108	1.69
DFT	-.28	.34	12	10,108	0.45
DMR	.25	.39	12	10,108	1.14
DMS	.22	.33	12	10,108	0.73

TABLE XVII.- Regression of the ORA Criterion  
on the Nine Predictors

Test	r	eta yx	No. of Columns	df	F
OTIS	-.15	.29	12	10,108	0.73
VR	-.01	.28	12	10,108	0.92
NA	-.11	.29	12	10,108	0.85
AR	-.18	.38	12	10,108	1.41
DSU	.35	.49	12	10,108	1.67
DMC	.02	.25	12	10,108	0.72
DFT	-.20	.31	12	10,108	0.67
DMR	.24	.37	12	10,108	0.99
DMS	.18	.35	12	10,108	1.11

TABLE XVIII.- Regression of the HIS Criterion  
on the Nine Predictors

Test	r	eta yx	No. of Columns	df	F
OTIS	-.09	.36	12	10,108	1.51
VR	.06	.25	12	10,108	0.68
NA	-.14	.21	12	10,108	0.28
AR	-.22	.37	12	10,108	1.11
DSU	.20	.38	12	10,108	1.32
DMC	-.03	.40	12	10,108	2.05*
DFT	-.14	.30	12	10,108	0.84
DMR	.23	.32	12	10,108	0.60
DMS	.14	.27	12	10,108	0.62

\*  $p \leq .05$

APPENDIX 2

FREQUENCY DISTRIBUTIONS FOR EACH VARIABLE AND SOME  
ASSOCIATED CHI SQUARE TESTS FOR GOODNESS OF FIT

TABLE XIX.- NA Chi Square Test

Scores	Original Frequencies		Regrouped Frequencies		fo - fe	(fo - fe) <sup>2</sup>	$\frac{(fo - fe)^2}{fe}$
	fo	fe	fo	fe			
40-41	3	3.9	3	3.9	0.90	0.81	0.21
38-39	8	9.0	8	9.0	- 1.00	1.00	0.11
36-37	26	16.0	26	16.0	-10.00	100.00	6.25
34-35	23	22.1	23	22.1	- 0.90	0.81	0.04
32-33	19	23.8	19	23.8	- 4.80	23.04	0.97
30-31	19	20.0	19	20.0	- 1.00	1.00	0.05
28-29	8	13.2	8	13.5	- 5.50	30.25	2.47
26-27	11	6.7	11	6.7	4.30	18.49	2.76
24-25	0	2.6	3	3.66	- 0.66	0.44	0.12
22-23	2	0.82					
20-21	0	0.20					
18-19	1	0.04					
$\Sigma$	120						12.98*

\* p ≤ .05, df = 6

TABLE XX.- HIS Chi Square Test

Scores	Original Frequencies		Regrouped Frequencies		fo - fe	(fo - fe) <sup>2</sup>	$\frac{(fo - fe)^2}{fe}$
	fo	fe	fo	fe			
92-97	2	1.8	2	1.8	0.2	0.04	0.02
86-91	3	5.5	3	5.5	- 2.5	6.25	1.14
80-85	10	13.6	10	13.6	- 3.6	12.96	0.95
74-79	38	23.3	38	23.3	14.7	216.10	9.27
68-73	22	27.9	22	27.9	- 5.9	34.81	1.25
62-67	26	24.0	26	24.0	2.0	4.00	0.17
56-61	8	14.4	8	14.4	- 6.4	40.96	2.84
50-55	7	6.6	7	6.6	0.4	0.16	0.02
44-49	0	2.0	4	2.5	1.5	2.25	0.90
38-43	3	0.4					
32-37	1	0.1					
$\Sigma$	120						16.56*

\*  $p \leq .05$ ,  $df = 6$

TABLE XXI.- VR Chi Square Test

Scores	Original Frequencies		Regrouped Frequencies		fo - fe	(fo - fe) <sup>2</sup>	$\frac{(fo - fe)^2}{fe}$
	fo	fe	fo	fe			
48-49	2	3.2	2	3.2	-1.2	1.44	0.45
46-47	9	6.4	9	6.4	2.6	6.76	1.06
44-45	16	10.7	16	10.7	5.3	28.09	2.63
42-43	16	15.8	16	15.8	0.2	0.04	0.003
40-41	18	18.9	18	18.9	-0.9	0.81	0.04
38-39	18	19.3	18	19.3	-1.3	1.69	0.09
36-37	13	16.8	13	16.8	-3.8	14.44	0.86
34-35	12	12.5	12	12.5	-0.5	0.25	0.02
32-33	7	7.9	7	7.9	-0.9	0.81	0.10
30-31	4	4.2	4	4.2	-0.2	0.04	0.01
28-29	4	1.7	5	2.4	2.6	6.76	2.82
26-27	1	0.7					
$\Sigma$	120						8.08*

\* N.S., df = 8

TABLE XXII.- AR Chi Square Test

Scores	Original Frequencies		Regrouped Frequencies		fo - fe	(fo - fe) <sup>2</sup>	$\frac{(fo - fe)^2}{fe}$
	fo	fe	fo	fe			
46-47	9	6.9	9	6.9	2.1	4.41	0.64
44-45	13	12.3	13	12.3	0.7	0.49	0.04
42-43	22	17.8	22	17.8	4.2	17.64	0.99
40-41	28	21.0	28	21.0	7.0	49.00	2.33
38-39	15	20.1	15	20.1	-5.1	26.01	1.29
36-37	12	16.7	12	16.7	-4.7	22.09	1.32
34-35	10	11.0	10	11.0	-1.0	1.00	0.09
32-33	6	5.9	6	5.9	0.1	0.01	0.00
30-31	3	2.6	5	3.9	1.1	1.21	0.31
28-29	0	0.9					
26-27	1	0.3					
24-25	0	0.1					
22-23	1	0.0					
$\Sigma$	120						7.01*

\* N.S., df = 6

TABLE XXIII.- DMS Chi Square Test

Scores	Original Frequencies		Regrouped Frequencies		fo - fe	(fo - fe) <sup>2</sup>	$\frac{(fo - fe)^2}{fe}$
	fo	fe	fo	fe			
12	1	0.8	2	2.9	-0.9	0.81	0.28
11	1	2.1					
10	5	4.5	5	4.5	0.5	0.25	0.06
9	8	8.2	8	8.2	-0.2	0.04	0.00
8	14	12.8	14	12.8	-1.2	1.44	0.11
7	17	16.9	17	16.9	-0.1	0.01	0.00
6	18	19.0	18	19.0	-1.0	1.00	0.05
5	16	18.3	16	18.3	-2.3	5.29	0.30
4	13	15.0	13	15.0	-2.0	4.00	0.27
3	16	10.1	16	10.1	5.9	34.81	3.45
2	5	6.2	5	6.2	-1.2	1.44	0.23
1	6	3.1	6	3.1	2.9	8.41	2.71
$\Sigma$	120						7.46*

\* N.S., df = 8

TABLE XXIV.- OTIS Frequency Distribution

Scores	Frequencies
72-74	2
69-71	6
66-68	10
63-65	10
60-62	16
57-59	25
54-56	19
51-53	18
48-50	8
45-47	4
42-44	2
N	120

TABLE XXV.- DSU Frequency Distribution

Scores	Frequencies
33-35	1
30-32	4
27-29	6
24-26	14
21-23	28
18-20	21
15-17	19
12-14	13
9-11	10
6-8	2
3-5	1
0-2	1
N	120

TABLE XXVI.- DMC Frequency Distribution

Scores	Frequencies
36-38	2
33-35	2
30-32	6
27-29	8
24-26	16
21-23	29
18-20	27
15-17	17
12-14	10
9-11	3
N	120

TABLE XXVII.- DFT Frequency Distribution

Scores	Frequencies
18-19	3
16-17	1
14-15	7
12-13	21
10-11	30
8-9	28
6-7	20
4-5	8
2-3	2
N	120

TABLE XXVIII.- DMR Frequency Distribution

Scores	Frequencies
22-23	1
20-21	0
18-19	5
16-17	11
14-15	14
12-13	25
10-11	26
8-9	21
6-7	10
4-5	5
2-3	2
N	120

TABLE XXIX.- GRA Frequency Distribution

Scores	Frequencies
92-96	1
87-91	2
82-86	13
77-81	10
72-76	19
67-71	26
62-66	13
57-61	13
52-56	18
47-51	1
42-46	2
37-41	1
32-36	1
N	120

TABLE XXX.- LIT Frequency Distribution

Scores	Frequencies
85-87	4
82-84	12
79-81	3
76-78	8
73-75	18
70-72	10
67-69	17
64-66	9
61-63	7
58-60	13
56-57	11
52-54	4
49-51	1
46-48	1
43-45	1
40-42	0
37-39	1
N	120

TABLE XXXI.- WRI Frequency Distribution

Scores	Frequencies
92-96	3
87-91	1
82-86	2
77-81	8
72-76	20
67-71	17
62-66	19
57-61	18
52-56	18
47-51	10
42-46	3
37-41	1
N	120

TABLE XXXII.- ORA Frequency Distribution

Scores	Frequencies
97-100	1
92-96	0
87-91	2
82-86	8
77-81	13
72-76	22
67-71	23
62-66	22
57-61	14
52-56	7
47-51	6
42-46	1
37-41	1
N	120

## APPENDIX 3

### ABSTRACT OF

#### The Relative Contribution of Some Divergent Thinking Tests to the Prediction of Academic Success<sup>1</sup>

At present, approximately twenty-five to fifty percent of the variance of a criterion of academic success can be predicted by cognitive tests. This situation has existed for the past thirty-five years. A possible reason for this plateau in prediction is Guilford's contention that only a limited number of human abilities have been represented in the test batteries employed. The Divergent Production tests of Guilford's Structure-of-Intellect Model appear to offer hitherto unmeasured factors of intellectual ability.

In this study, the cross-validated differences in academic prediction between a standard aptitude battery and the same battery with some Divergent Thinking tests added were investigated.

The null hypothesis of no significant differences in cross-validated predictive validity between the two batteries was accepted. Since many of the prediction equations did not cross-validate, a Discriminant Analysis was performed which

---

<sup>1</sup> Walter K. Noë, master's thesis presented to the Faculty of Psychology of the University of Ottawa, Ontario, 1971, vi-79 p.

revealed a significant difference between the Validation and Cross-Validation groups on the measures used. Therefore, the same subjects were reassigned to two new groups in such a manner that the test performance of those groups was equated. However, the results obtained were very similar to those obtained with the original two groups. Possible reasons for the failure to cross-validate are discussed, centering around the limitations of the technique itself and the reliability of the measures used.