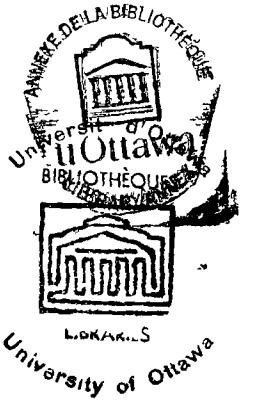
DOES EITHER A MODERATOR MODEL OR A MULTIPLICATIVE MODEL IMPROVE PREDICTION OVER THE USUAL MULTIPLE REGRESSION MODEL IN THE CONTEXT OF AN INTENSIVE FRENCH LANGUAGE TRAINING PROGRAM?

by: Charles E. McInnis



Thesis presented to the School of Graduate Studies as partial fulfilment of the requirements for the degree of Doctor of Philosophy

Faculty of Psychology
UNIVERSITY OF OTTAWA
CANADA, 1972

ACKNOWLEDGMENTS

This thesis was prepared under the supervision of Professor Lawrence T. Dayhaw, Ph.D., Vice-Dean of the Faculty of Psychology. The writer gratefully acknowledges his advice, direction, and interest in the research.

The writer further wishes to thank Professor Olgierd Porebski, Ph.D. for his advice and assistance, in particular for making available and interpretable the computer programs required for conduct of the research. The co-operation of the staff in the Canadian Public Service Language Schools, in particular Mr. Mabela Rongongo, Mr. Urbain Morin and Mr. Marc Robillard, was appreciated for enlisting the support of all personnel who participated in and made possible the completion of the research.

Special gratitude is also extended to the Consulting Psychologists Press, Inc., publishers of the California Psychological Inventory, for their kind permission to reproduce the three copyrighted scales for research purposes, and also to Dr. W. Lambert of McGill University for permission to use the Integrative Orientation measure.

CURRICULUM STUDIORUM

Charles E. McInnis was born July 12, 1924 in Antigonish, Nova Scotia. He received his Bachelor of Arts Degree from St. Francis Xavier's University, Antigonish, Nova Scotia in 1946. After a period of service in the Canadian Forces, he received his Master of Arts degree in Psychology from the University of Ottawa, Ottawa, Ontario in 1967. The title of his thesis was: "The Effect of Peer Rating Treatments on Group Morale", in Canadian Psychologist, Vol. 11, No. 3, July 1970, p. 281-288.

TABLE OF CONTENTS

LIST OF TABLES

Table page

Table page

LIST OF FIGURES

Figure page

INTRODUCTION

It has been shown, using classical validation models, that predictive validities rarely exceed .50 for training criteria and .35 for performance criteria. Related to this fact is the general observation that multiple correlations rarely increase significantly after four or five tests have been used in the regression equation. It has generally been conceded that the same limited number of fundamental abilities are tapped by many different tests.

In an attempt to up-date the classical validation model and to improve prediction, many different researchers have developed various strategies. The nomenclature applied to these strategies has varied, i.e., moderator variable, modifier variable, population control variable, sub-grouping variable, predictability variable, etc..., but the objective is similar: to increase the predictive ceiling of psychological tests.

In the area of personality testing in particular there has been a proliferation in the development of tests but in no other field have the results been as disappointing. This has led some researchers to state that a moratorium should be placed on the development of new tests until better use is made of some of the tests currently available. There has been the suggestion that tests which purport to measure motivation and interest in particular may function better as

moderators than as predictors in selection research models. The research literature also indicates that the combined effects of motivation and ability on performance may be interactive and multiplicative rather than additive.

Many of the researchers who have attempted to investigate these problems have been criticized in that either the predictor or the criterion variables used were not sufficiently reliable. Since much research has been conducted in the area of intensive foreign language learning and relevant measures have been developed that are demonstrated to possess a high level of reliability, it was considered that such an area would lend itself to the testing of differential prediction models.

The purpose of this research, therefore, is primarily to investigate whether relevant measures of motivation and of orientation might function better as moderators than as predictors in the context of a Canadian Public Service Intensive French Language Training Program. A secondary aim was to test whether the effects of ability and motivation or orientation on performance in language learning are really additive or whether they are multiplicative and interactive.

Chapter 1 will review the literature on moderator and interactive models generally, and also consider the variables found relevant in the prediction of success in foreign language learning. The Chapter will conclude with the null hypotheses for the research. Chapter 2 will present the experimental

design while Chapter 3 will present the results and a discussion. The Chapter will conclude with a summary of findings and suggestions for further research.

C H A P T E R 1

REVIEW OF THE LITERATURE AND STATEMENT OF THE PROBLEM

The research literature in the area of personnel selection attests to the concern of psychologists for the low validity of currently used psychological tests. Over several decades of effort two facts have not changed. First, validity coefficients rarely exceed .50 and show no time trend toward improvement. Second, training criteria are more predictable than job performance criteria.

These generalizations are confirmed by Ghiselli[1] who published all the reasonably solid validity coefficients he could find, classifying them in terms of both occupation predicted and kinds of tests used. On the basis of averaging the validity coefficients of the best test for each job, Ghiselli concludes:

> Taking all jobs as a whole... it can be said that by and large the maximal power of tests to predict success in training is in the order of .50, and to predict success on the job itself is of the order of .35....[2]

In other words, such validity coefficients account for 25% of the training criterion variance and 12% of the performance

---

1 Edwin E. Ghiselli, The Validity of Occupational Tests, New York, John Wiley and Sons, Inc., 1966, 155p.

2 Idem, ibid., p.125.

criterion variance. The range of average validity coefficients reported was .27 to .59 for training criteria and .16 to .46 for proficiency criteria.

Psychologists concerned about the apparent prediction ceiling of psychological tests have devoted much effort and attention to the possibility that significant increases in predictive validity may reward the researcher who discovers a method to maximize test validity without restricting himself to the widely used linear methods. A variety of approaches to the problems of nonlinear combinations of validity data are found in the literature. In identifying their approaches to the problem, different researchers have utilized terms such as: pattern analysis, profile analysis, configural scoring, the use of moderator and supressor variables, addend scoring, interaction analysis, and differential predictability.

In the attempts to update the theoretical model underlying test validation and selection research the concept of a moderator model originally proposed by Saunders[3] became a promising technique. In the usual moderator model, the validity of a predictor is considered separately at different levels of a second or of a third variable. Another approach suggested

---

3 D.R. Saunders, "Moderator Variables in Prediction", in Educational and Psychological Measurement, Vol. 16, 1956, p. 209-222.

by Vroom[4], Lawler[5], specifies that performance can be thought of as a multiplicative function of motivation and ability [$P = f(M \cdot A)$]. Attention to this latter model demonstrates that performance has two components, namely, motivation (M) and ability (A), and emphasizes that unless both measures of M and A are obtained, errors are likely to occur in evaluating the utility of ability tests. According to this model unless motivation is high, ability will not be related to performance. It would also follow from this model that unless ability is high, motivation will not be related to performance.

The general models enumerated above will be tested empirically in a setting where subjects are engaged in intensive French Language training. The area of French language learning has been selected for a number of reasons:

1. model testing requires a large sample of subjects and intensive French language training is being conducted on a large scale in numerous settings;
2. much relevant research has been conducted in the area of language learning; and
3. standardized measures have been developed and utilized which purport to measure language ability, orientation to learning the language, motivational aspects of language learning, and achievement in language learning.

---

4 V.H. Vroom, Work and Motivation, John Wiley and Sons Ltd., New York, 1964, p. 331.

5 E.E. Lawler, "The Multi-Trait Multirater Approach to Measuring Managerial Job Performance", in Journal of Applied Psychology, Vol. 51, No. 5, October 1967, p. 369-381.

REVIEW OF THE LITERATURE AND STATEMENT OF THE PROBLEM

Prior to defining the problem in detail and stating the hypotheses which will be tested in this research, a review of the literature relevant to the research will be presented. The review will be organized in four sections as follows:

1. Research on Validity of Moderator Models-General;
2. Research on General Validity of Multiplicative Model $[p = f(M \cdot A)]$;
3. Research on Use of Personality and Motivation Measures as EITHER Moderators OR Predictors in Selection Research Models;
4. Research on Utility of Measures of Aptitudes, Orientation, Motivation and Performance in Language Learning:
    a. McGill Studies;
    b. Peace Corps Studies;
    c. Studies Using Achievement Motivation Scales of the California Psychological Inventory as a Measure of Achievement Motivation

1. Research on Moderator Models-General

In a survey of the historical development and current status of research on moderator variables, Banas proposed this following definition of a moderator. It is:

> The general term to refer to all variables, quantitative and qualitative, which improve the usefulness of a predictor by isolating subgroups of individuals for whom a predictor or set of regression weights are especially appropriate.[6]

---

6 P.M. Banas, An Investigation of Transituational Moderators, Unpublished Doctoral Dissertation, University of Minnesota, 1964.

From the above definition, validity of a test or battery for predicting a given criterion (criteria) may vary among subgroups differing in personal characteristics. The validity coefficient when computed in a total group may be too low to be of practical significance. When computed for subsets of individuals, validity may be high in one subset and low in another. In the language of decision theory the test could be used to make decisions for the first group but not for the second group.

Approaches to the development of moderator variables have been offered by many different researchers. In an evaluation of moderator models up to 1964, Banas[7] concluded that Ghiselli's[8,9] D-Score moderator model had the greatest appeal due to the directness of application and empirical support. This model enables the researcher to analyze data, in order to identify variables that correlate with absolute differences (D's) between standard scores on the predictor and standard scores on the criterion. Other approaches evaluated by Banas

---

7 Idem, ibid.

8 Edwin E. Ghiselli, "Differentiation of Individuals According to Their Predictability", in Journal of Applied Psychology, Vol. 40, 1956, p. 374-377.

9 --------, "Differentiation of Tests in Terms of the Accuracy with which They Predict for a Given Individual", in Educational and Psychological Measurement, Vol. XX, No. 4, 1960, p. 675-684.

were models offered by Lykken[10], Toops[11], Rimland[12] and French[13].

From the research on moderator variables it appeared that moderator variables were being found - but only as often as chance permitted. When Ghiselli's[14,15] D-Score moderator model failed to produce positive results, Dunnette[16], and Hobert and Dunnette[17] continued their efforts and developed an alternative method called Quadrant Analysis (Q.A.) for discovering correlates of individual differences in predictability. The Q.A. method identifies subsets of subjects overpredicted and

---

10 D.T. Lykken, "A Method of Actuarial Pattern Analysis", in Psychological Bulletin, Vol. 53, 1956, p. 102-107.

11 H.A. Toops, "A Research Utopia in Industrial Psychology", in Personnel Psychology, Vol. 12, 1959, p. 189-225.

12 B. Rimland, "Multidimensional Scatterplotting: A Graphic Approach to Profile Analysis", in Journal of Applied Psychology, Vol. 44, 1960, p. 404-406.

13 J.W.W. French, A Machine Search for Moderator Variables in Massive Data, Princeton, New Jersey: Educational Testing Service, 1961.

14 Ghiselli, Op. Cit., 1956, p. 374-377.

15 --------, Op. Cit., 1960, p. 675-684.

16 Marvin D. Dunnette, "A Modified Model for Test Validation and Selection Research", in Journal of Applied Psychology, Vol. 47, 1963, p. 317-323.

17 Robert Hobert and Marvin D. Dunnette, "Development of Moderator Variables to Enhance the Prediction of Managerial Effectiveness", in Journal of Applied Psychology, Vol. 51, 1967, p. 50-64.

underpredicted by dividing at the median on both predictor and criterion scores. Using this method, two moderator scales were empirically developed to identify and eliminate the "off-quadrant" cases. The elimination of these "off-quadrant" cases resulted in improved prediction and reduced criterion overlap for the remainder of the sample.

A subsequent discussion by Abrahams[18] points out a serious methodological problem with the Q.A. model. To the extent of the relationship between the predictor and the criterion, cases above and below the criterion median within each predictor category are not equivalent on predictor scores. Because of these differences, items or scores that differentiate between comparison groups may merely reflect the predictor composite differences. It is further suggested by Abrahams[19] that perhaps this difficulty could be overcome by matching comparison groups on predictor-composite scores. The suggestion previously made by Hobert and Dunnette[20] that moderators outside the original predictor set may be more efficient should perhaps be followed in any future application of the Q.A. approach.

---

18 Norman N. Abrahams, "Off-Quadrant Comment", in Journal of Applied Psychology, Vol. 53, No. 1, 1969, p.66-68.

19 Idem, ibid., p.68.

20 Hobert and Dunnette, Op. Cit., p.60.

A modification of the Q.A. model is suggested by McNemar[21] and illustrated by way of an idealized set of data. It is shown by McNemar[22] that in terms of increased validity coefficients, decreased percentage of overlap and increased hit rate, the modification leads to increased improvement over the original Hobert and Dunnette Q.A. method. Then, McNemar[23] points out the seriously misleading nature of this search for moderators for both original and modified versions of the Q.A. model.

Criticisms levelled at existent moderator models have not slowed down attempts to find and apply variables that moderate the validity of psychological tests. Much of the impetus for this research stems from the problem of selection and discrimination in employment practices. The 1964 Civil Rights Act in the United States makes the question of discrimination by tests on race, color, religion, national origin, and sex, a legal as well as an ethical question. The question according to Quion[24] is to determine if a test does discriminate against

---

21 Quinn McNemar, "Moderation of a Moderator Technique", in Journal of Applied Psychology, Vol. 53, No. 1, 1969, p. 69-72.

22 Idem, ibid., p. 69.

23 Idem, ibid., p. 69.

24 R.M. Quion, Personnel Testing, New York: McGraw-Hill, 1965, p. 492-493.

a minority group, and if so, develop a means to eliminate discrimination.

In the search for moderator variables, race has been used most frequently. In a recent report, Bartlett and O'Leary[25] summarize much of the research on the use of race as a moderator variable. They present eleven different models which represent possible relationships which may exist when heterogeneous groups are combined in making predictions. Research was documented for ten of the eleven models presented. The most unusual situation cited was a research reported by Lopez[26] where there was no significant differences in means on selection tests between two racial groups; yet Lopez[27] reported significant validities in the opposite direction for the two groups. Combining the two groups in this case would result in zero validity for the tests. The proper use of the tests in personnel selection would mean hiring whites who score high and non-whites who score low. In short, when the nature of heterogeneity can be identified such as different sexes, races or

---

25 C.J. Bartlett and B.S. O'Leary, "A Differential Prediction Model to Moderate the Effects of Heterogeneous Groups in Personnel Selection and Classification", in Personnel Psychology, Vol. 22, 1969, p. 1-17.

26 F.M. Lopez "Current Problems in Test Performance of Job Applicants", in Personnel Psychology, Vol. XIX, 1966, p. 10-17.

27 Idem, ibid., p. 10-17.

socio-economic backgrounds, better prediction is possible by validating the tests separately for each identifiable group. As Krug has stated:

> In validating a selection technique it is always desirable to study as many separate groups as possible in preference to one large heterogeneous group.[28]

Apart from race, Dawis et al[29] found that satisfaction moderated the relationship between ability test scores and performance for male and female assemblers. Similarily, Stricker[30] found that measures of compulsivity moderated the prediction of college grades for engineering students but not for those in liberal arts. In another study, Baker[31] applied a bio-data predictor to a large sample of employees of a major oil company. After factor analysis of his data, he used factor scores to predict both salary level and potential salary level; each factor was also tested for moderator effects

---

28 R.E. Krug, "Some Suggested Approaches for Test Development and Measurement", in Personnel Psychology, Vol. XIX, 1966, p. 24-34.

29 R.V. Dawis, E. Bettz, L.H. Lofquist, D.T. Weiss Paper Read at 75th Annual American Psychological Association Convention, Vol. 2, 1967, p. 369-370.

30 L.J. Stricker, "Compulsivity as a Moderator Variable: A Replication and Extension", in Journal of Applied Psychology, Vol. 30, No. 4, 1966, p. 331-335.

31 B.L. Baker, The Use of Biographical Factors to Moderate Prediction and Predict Salary Level, Unpublished Doctoral Dissertation, Purdue University, Lafayette, Indiana, 1967.

by dividing its distribution into upper and lower halves to determine if prediction in one group of Ss was superior to prediction in the other group. In the same study, the two best moderators were selected and their combined effects were observed in a fourfold table. On cross-validation, it was found that the moderator approach had an inconclusive effect on overall prediction but did enhance prediction within certain sub-groups.

Possibly some encouragement on the advisability of continuing to pursue the moderator variable approach in personnel selection research is an evaluation by Owens and Jewell as follows:

> It must be noted that the new look at prediction is a differential look. Correlations in some segments of the sample are increased and those in other segments are constant or decreased; the total may remain essentially unchanged. Moderators tend to be specific and to require large N's for their development. Methods of subgrouping, although they do have inferential utility, are numerous and plagued with a variety of inexactitudes and uncertainties. If the subgrouping is accomplished to enhance prediction within subsets, restriction in range, relative instability of betas, and similar methodological problems arise; one must also hope that the subgroup within which prediction is enhanced is one of interest and concern. Some of the problems are lost if attention is given only to the simple fact of subgroup membership and to the differential affinities of subsets for criteria. In any extent, the analytical search for enhanced differential validities is clearly a promising move

> in the right direction. Given some of the indicated problems, however, the form it may ultimately assume is only beginning to emerge[32].

A more recent theoretical evaluation of the use of moderator variables in prediction is presented by Zedeck[33]. Alternative concepts of moderator variables are defined and categorized. The statistical and methodological problems involved in different approaches are elaborated. In the interpretation and discussion of the findings of this research, reference will be made to this article and an attempt will be made to deal with some of the problems outlined.

2. Research on General Validity of a Multiplicative or Interactive Model $[p = f(M \cdot A)]$

A number of investigations in very different settings support the validity of the multiplicative or interactive model. In a study with basic airmen in the United States Air Force (USAF), French[34] classified subjects according to

---

32 W.H. Owens and D.O. Jewell, "Personnel Selection", in Annual Review of Psychology, Vol. 20, 1969, p. 434-436.

33 Sheldon Zedeck, "Problems With the Use of Moderator Variables", in Psychological Bulletin, 1971, Vol. 76, No. 4, p. 295-310.

34 E.G. French, "Effects of Interaction of Achievement Motivation and Intelligence on Problem-Solving Success", in American Psychologist, Vol. 12, 1957, p. 339-400.

three intelligence and two achievement motivation levels. A significant F was found for the interaction between motivation and intelligence. The means on problem solving for the three high motivation groups were separated according to intelligence level while those of the three low motivation groups were not.

Similar results were found by Fleishman[35] when he experimentally manipulated the motivation of USAF enlisted men in learning a perceptual motor skill. The main effects, ability level and motivation level, as well as their interaction effect were found to be significant beyond the .01 level. The interpretation of the significant interaction effect was that the motivation treatment made no difference in the performance of the low-ability subjects but made a difference in the performance of the high-ability subjects.

In three US state government settings, Lawler[36] measured the degree to which middle and low level managers felt their pay to be contingent upon their job performance. Job performance and ability measures were obtained from superiors' rankings and

---

35 E.A. Fleishman, "A Relationship Between Incentive Motivation and Ability Level in Psychomotor Performance", in Journal of Experimental Psychology, Vol. 56, 1958, p. 78-81.

36 Edward E. Lawler III, "Ability as a Moderator of the Relationship Between Job Attitudes and Job Performance", in Personnel Psychology, Vol. 19, 1966, p. 153-164.

from self-ratings. The results of this study supported the hypothesis that $[p = f(M \cdot A)]$. For managers categorized as low in ability, no significant difference was found between high and low contingency groups in terms of their job performance. For the managers who were high in ability, a significant difference was found between the high and low contingency groups. The data gathered for self-ratings, as well as the data accumulated from superiors' rankings, fit the multiplicative model.

A study by Galbraith and Cummings[37] in a heavy equipment manufactures plant in Indiana was designed to operationalize and test Vroom's[38] model. Using a modified analysis of variance procedure, results yielded support for hypothesized interactive effects between measures of motivation and ability.

In an appraisal of Galbraith and Cummings' research, Hinrichs states as follows:

> Galbraith and Cummings attempt to test Vroom's suggestion that performance is a multiplicative function of motivation and ability. They operationalized motivation as a function of valences of job outcome (money, fringe benefits, promotion, managerial supportiveness, and group acceptance) and the perceived instrumentality of performance for each factor. Ability was operationalized as

---

37 Joy Galbraith, and L.L. Cummings, "An Empirical Investigation of the Motivational Determinants of Task Performance: Interactive Effects Between Instrumentality-Valence and Motivation-Ability", in Organizational Behavior and Human Performance, Vol. 2, 1967, p. 237-257.

38 Vroom, Op. Cit., 1964, p. 331.

> time on the job. Their analysis suggested that ability (tenure) does interact multiplicatively with motivation in determining the relationship with performance, but the data are far from clear cut. Galbraith and Cummings point to problems of restricted variance in their measures, but it is also clear that any model such as this suffers from the multiplicative effects of error variance. Also more appropriate measures of ability would have been desirable. But their study illustrates the type of comprehensive model testing that is needed and that should be attempted within more heterogeneous samples with more reliable measures.[39]

It would appear from the literature review on moderator and interactive or multiplicative models that there is considerable support for the need for additional research in varied settings aimed at the extension and testing of such models. The third section of this literature review will present research relevant to the use of measures of motivation and attitude as moderators vs predictors in selection research models.

3. Research on Use of Personality and Motivation Measures as EITHER Moderators OR Predictors in Selection Research Models - Specific

It is suggested by Ghiselli[40] that at least one logical place to look for moderator variables in the future is in the

---

39 John R. Hinrichs, "Psychology of Men at Work", in Annual Review of Psychology, Vol. 21, 1970, p. 535.

40 Edwin E. Ghiselli, "Moderating Effects and Differential Validity", in Journal of Applied Psychology, Vol. 42, 1963, p. 81-86.

direction of measures of motivation. A typical approach in the past has been to measure achievement motivation through projective techniques such as Murray's[41] Thematic Apperception Test, French's[42] "Test of Motivation" disguised as a "Test of Insight", or use of self-report measures. Guion and Gottier[43] attempt to report in summary form the nature of -- and results of -- a sample of research studies featuring different personality measures. Two major sources of published research reports were summarized: The Journal of Applied Psychology and Personnel Psychology. Twelve volumes of each of these journals were examined, from 1952-1963, and the low validity of available tests of personality and motivation is emphasized. Guion and Gottier further report Guion's theme that:

> A growing body of research literature suggests that motivation variables may serve better in the role of moderator variables than as predictors.[44]

---

41 H.A. Murray, Thematic Apperception Test Manual, Cambridge: Harvard University Press, 1943, 20p.

42 E.G. French, Development of Measures of Complex Motivation, Research Report AFPTRC-TN-56-48, Air Force Personnel and Training Research Centre, San Antonio, Texas, April 1956, 10p.

43 R.M. Guion and R.F. Gottier, "Validity of Personality Measures in Personnel Selection", in Personnel Psychology, Vol. 18, 1965, p. 135-164.

44 Idem, ibid., p. 158.

The next stage in the literature review will encompass predictive studies in the specific area of intensive language learning. This review will facilitate the selection of vari ables that have been found to be useful, reliable, and valid in the context of intensive foreign language training programs.

4. Research on Utility of Measures of Orientation, Motivation, Aptitude and Performance in Language Learning Specific

As a necessary step in the selection of relevant variables to permit empirical testing of the models enumerated above in a language setting, particular attention was directed to a series of studies conducted by Gardner and Lambert[45], Lambert et al[46,47,48] of McGill University and Peace Corps

---

45 R.C. Gardner and W.E. Lambert, "Motivational Vari ables in Second Language Acquisition", in Canadian Journal of Psychology, Vol. 13, 1959, p. 266-272.

46 W.E. Lambert, R.C. Hodgson, R.C. Gardner and S. Fillenbaum, "Evaluational Reactions to Spoken Languages", in Journal of Abnormal and Social Psychology, Vol. 60, 1960, p. 44-51.

47 W.E. Lambert, R.C. Gardner, H.C. Barek and K. Tunstall, "Attitudinal and Cognitive Aspects of Intensive Study of a Second Language", in Journal of Abnormal and Social Psychology, Vol. 66, 1963, p. 358-368.

48 W.E. Lambert, R.C. Gardner, R. Olton and K. Tunstall, A Study of the Roles of Attitude and Motivation in Second Language Learning, Unpublished Manuscript, McGill University (undated), p. 1-159.

studies reported on by Carroll[49,50], Krug[51], and Hobbs[52]. On the basis of factor analysis in these researches two independent dimensions or factors were found to underlie the development of skill in learning a second language, an intellectual capacity (aptitude-intelligence dimension), and appropriate orientation or attitude towards the other language group coupled with a motivation to learn the language (orientation-motivation dimension). Furthermore, the McGill studies indicated that Ss with an "integrative orientation" performed better in acquiring a second language than Ss with an "instrumental orientation". "Integrative" and "Instrumental" orientation are explained in these terms:

> The orientation takes an "instrumental" form if the purpose of language study reflects the more utilitarian value of linguistic achievement such as getting ahead in ones work, and an "integrative" form if the student is oriented towards learning

---

49 J.B. Carroll, "The Prediction of Success in Intensive Language Training", in R. Glaser (Ed.), Training Research and Education, Pittsburgh: University of Pittsburgh Press, 1962, p. 87-136.

50 --------, A Parametric Study of Language Training in the Peace Corps, Fiscal Report for Contract PC - (W) 226 with the Peace Corps, 1966, 146p.

51 R.E. Krug, An Analysis of Eighteen Peace Corps Projects, Pittsburgh: American Istitute for Research, 1962, 52p.

52 N.A. Hobbs, "A Psychologist in the Peace Corps", in American Psychologist, Vol. 18, (1), 1963, p. 47-55.

> more about the other cultural community, as if he were eager to be a potential member of the other group.[53]

Theoretical attempts to explain how a child learns his first language have emphasized a particular type of motivation. A theory presented by Mowrer[54] suggests that language acquisition is motivated by a desire to be like the valued members of a family, and later of the whole linguistic community. This interpretation would seem to be similar to "integrative orientation" as explained by Lambert[55].

Other measures used in the McGill studies to measure attitude and the achievement motive were not considered appropriate for this research. As an example, Lambert[56] developed a scale to measure "intensity of motivation to learn French" but a close scrutiny of the items would suggest that the scale would be more appropriate for research with a high school population. A large number of Ss engaged in the Public Service Language Program have been out of high school for over 20 years.

---

53 Lambert et al, Op. Cit., 1963, p. 358.

54 O.H. Mowrer, Learning Theory and Personality Dynamics, New York: Ronald Publishing Co., 1950, V111 - 776 p.

55 Lambert et al, Op. Cit., 1963, p. 358.

56 Gardner et Lambert, Op. Cit., 1959, p. 268.

The three achievement scales of Gough's[57] California Psychological Inventory (CPI) provide a means towards the assessment of the contribution afforded motivational, non-intellectual factors in academic achievement. Research studies by Gough[58], Holland[59], Maxwell[60], Rosenberg et al[61] have demonstrated empirically its utility.

Of particular relevance to this research is a study conducted by Datel et al[62] in US Army Security Agency Foreign Language Candidate Training Program. The three achievement

---

57 H.G. Gough, Manual for the California Psychological Inventory, Palo Alto, California: Counsulting Psychologists Press, 1957, 22p.

58 --------, A Comparison of College Dropouts and College Graduates, Paper Read at the Annual Meetings of the American Personnel and Guidance Association, 1961.

59 John L. Holland, "The Prediction of College Grades from the California Psychological Inventory and the Scholastic Aptitude Test", in Journal of Educational Psychology, Vol. 50, No. 4, 1959, p. 135-142.

60 Martha J. Maxwell, "An Analysis of the California Psychological Inventory and the American Council on Education Psychological Test as Predictors of Success in Different College Curricula", in American Psychologist, Vol. XV, 1960, p. 425.

61 L.A. Rosenberg, T.B. McHenry, Anna M. Rosenberg, and R.C. Nichols, The California Psychological Inventory as a Potential Screening Device in an Academic Setting, Paper Read at the Annual Mettings of the American Psychological Association, New York City, September 1961.

62 William E. Datel, Forrest D. Hall, and Charles P. Rufe, "Measurement of Achievement Motivation in Army Security Agency Foreign Language Candidates", in Educational and Psychological Measurement, Vol. XXV, No. 2, 1965, p. 539-545.

scales of the CPI, Achievement via Independence (Ai), Intellectual Efficiency (Ie), and Achievement via Conformance (Ac), which are thought to measure academic potential and the achievement motive, were chosen out of the total pool of 480 items making up the CPI. The 105 items composing the three scales were mimeographed into test booklets headed by the usual "agree-disagree" instructions. The scales were administered to 290 incoming students who subsequently went on to study any one of 18 different foreign languages, for a period of approximately 47 weeks. Two of the three CPI scales, Ai and Ie, had a significant but modest correlation with final grades. The authors point out the high degree of pre-selection for the sample studied in that on the ability measure the mean score for the group is two standard deviations above the mean of the standardization sample. In the same study using a "completion" "dropout" dichotomy the Ai and Ie scales discriminated significantly between the two groups whereas the more traditional predictors of achievement (general Ability Test, and US Army Aptitude Test Battery) failed to differ in the split sample. It was concluded from the study that:

> students who successfully complete a foreign language course tend to obtain higher scores on Ai and Ie, (and likewise on the Ac) than do students

> who for some reason or other fail to complete successfully the course.[63]

In the assessment of aptitude for foreign language learning, the question of using intelligence tests has often been raised. One of the purposes of research leading to the development of Carroll-Sapon's[64] Modern Language Aptitude Test (MLAT) was the investigation of which factors of ability should be included in a language prognosis test and which could safely be omitted. Some of the conclusions to which this research led are recorded in a report by Carroll[65]. In a subsequent report Carroll states as follows:

> .... facility in learning to speak and understand a foreign language is ... relatively independent of the traits ordinarily included under intelligence ....[66]

It seems rather safe to conclude that the MLAT measures a rather special collection of abilities probably many of them not being represented at all in the typical intelligence test.

---

63 Idem, ibid., p. 543.

64 John B. Carroll, and Stanley M. Sapon, Modern Language Aptitude Test, New York: The Psychological Corporation, 1965, 12p.

65 John B. Carroll, "A Factor Analysis of two Foreign Language Aptitude Batteries", in Journal of General Psychology, Vol. 59, 1958, p. 3-19.

66 Carroll and Sapon, Op. Cit., 1962, p. 89.

Scores on the MLAT have been found to be highly predictive of success in foreign language study, particularily in intensive language courses[67]. The Research Division of the Peace Corps has conducted or sponsored numerous studies showing satisfactory validity of the test in predicting success in Peace Corps language training programs[68]. On the validity of the MLAT in predicting success not only in the training program but also in the operational setting Hobbs points out that the test battery:

> .... holds up reasonably well not only for countries where a new language must be learned but for others as well, suggesting the importance of some general ability -- to -- learn factor.[69]

Indeed several parts of Carroll-Sapon's[70] MLAT are explicitly learning tasks, particularily sub-tests, 1, 11 and V.

Last, but perhaps most important in the review of research literature directed to the selection of relevant measures to test the moderator and multiplicative models suggested by respectively Guion[71] and Vroom[72], is the adoption

---

67 Idem, ibid., p. 87-136.

68 Krug, Op. Cit., 1962, 52p.

69 Hobbs, Op. Cit., 1963, p. 47-55.

70 Carroll and Sapon, Op. Cit., 1965, p. 3.

71 Guion and Gottier, Op. Cit., p. 158.

72 Vroom, Op. Cit., 1964, p. 331.

of reliable and valid criterion measures. The most appropriate measures currently available and widely used in Peace Corps studies[73] would appear to be the MLA -- Cooperative Foreign Language Tests -- French version[74]. Two parallel forms of the test are available for each of two levels of achievement. The forms are comparable in content coverage and difficulty, level L designed for the first and second levels in secondary school and college, and level M corresponding to the third and fourth years of language learning in secondary school or college. The adoption of the MLA as the criterion measure would provide a standardized language proficiency test that can be applied uniformly to all Ss engaged in the intensive French language training program, and would ensure high criterion reliability.

This concludes the analysis of the research literature which influenced the development of the research and the choice of measuring instruments. In the following section, a brief summary details the main points considered in the literature review.

---

73 Carroll, Op. Cit., 1962, p. 87-136.

74 Manual, MLA -- Cooperative Foreign Language Tests -- French Versions, Educational Testing Service, 1969, p. 20-21.

5. Summary Statement on Literature Review and Statistical Considerations.

The preceding sections presented briefly the review of the literature on moderator models and interactive models generally. Since the models are to be tested empirically in the setting of an intensive French language training program, a review was also presented on variables found to be predictive in the area of foreign language training. These variables will be described in more detail in the chapter on Experimental Design.

Before stating the problems to be investigated in the research some preview of the statistical procedures employed in prediction type studies was considered to be mandatory. As an example, Darlington[75] criticizes a number of common practices and beliefs concerning multiple regression, and several paradoxical properties of the method are emphasized. In a discussion of statistical criteria for the selection of predictor variables, Darlington[76] defends stepwise regression models and extensions thereof, and shows that other selection procedures are inferior to the stepwise method. In another

---

75 Richard B. Darlington, "Multiple Regression in Psychological Research and Practice", in Psychological Bulletin Vol. 69, No. 3, 1968, p. 161-182.

76 Idem, ibid., p. 180.

study Herzberg[77], using both theoretical and Educational Testing Service data, studied several methods of multiple regression and found that of two reduced rank methods, the principal predictors method had generally superior or equal cross-validities to the principal components method. Herzberg suggests that:

> This result is similar to Burket's[78] (1964) finding that the principal components correlating greatest with the criterion do not validate as well as the largest principal components. It appears to be an advantage to select linear combinations of the predictors independently of criterion information in order to maximize cross-validity.[79]

On the other hand, Cooley and Lohnes[80] have reservations about the stepwise procedure and the test of significance for the standard partial-regression coefficient it utilizes in selecting the variables to be added or dropped. They state:

> If we remind ourselves that a multiple regression is really a univariate model, since only the criterion variable is treated as subject to errors, it follows that any

---

77 Paul A. Herzberg, "Parameters of Cross-Validation", in Supplement to Psychometrika, 1969, p. 1-70.

78 G.R. Burket, "A Study of Reduced Rank Models for Multiple Prediction", in Psychometric Monographs, No. 12, 1964.

79 Herzberg, Op. Cit., p. 66.

80 W.W. Cooley, and P.R. Lohnes, Multivariate Procedures for the Behavioral Sciences, New York, John Wiley and Sons, 1962, p. 35.

> effort to generalize from sample to population is open to serious danger of capitalization on chance, particularly if it involves selecting some predictors and discarding others.[81]

In a more recent study, McCornack[82] compares three multiple prediction techniques in terms of shrinkage in cross-validation. An iterative method was found to select predictors with slightly, but consistently, higher cross-validities than the popularly used stepwise method. This was particularily true for sample sizes under 100 and its superiority seemed to increase as more predictors were added. A gradient method was found to be comparable to the stepwise method only in the larger samples and for the largest predictor subsets.

It is further stated by McCornack[83] that no multiple regression method is known that will select a subset of predictors which will be optimum on cross-validation. This suggests an empirical comparison. In this study the decision was made to test empirically two models, a "stepwise" model and a "full" multiple regression model in a double cross-

---

81 Idem, ibid., p. 35.

82 Robert L. McCornack, "A Comparison of Three Predictor Selection Techniques in Multiple Regression" in Psychometrika, Vol. 35, No. 2, June 1970, p. 257-271.

83 Idem, ibid., p. 257.

validation design[84]. By means of cross-validation, it is possible to estimate the correlation in new samples using only data in a (divided) original sample. The choice of the "full" or "unrestricted" model over an iterative model was determined largely from the fact that in the area of intensive language learning it was possible from published research to select a relatively small number of relevant predictor variables.

6. Statement of the Problem (s)

This section of the chapter will state the problems to be investigated in the research and will terminate with a rewording of the problems in null hypotheses form. The problems are operationalized as follows:

Problem 1. Compare cross-validated Multiple R's for Stepwise and Full Multiple Regression Models.

Problem 2. After determining empirically the regression model that yields the best double cross-validation results, using that model to test:

a) If Achievement Motivation and Integrative Orientation contribute jointly to prediction over and above the aptitude predictors.

---

84 C.I. Mosier, "Problems and Designs of Cross-Validation", in Educational and Psychological Measurement, Vol. 11, 1951, p. 5-11.

Problem 2. (Cont'd)

b) If either Achievement Motivation or Integrative Orientation contribute singly to prediction over and above the aptitude predictors.

For the next sequence of problems the term "Validity Coefficient" will be used to demonstrate the relationship between the aptitude predictors (sub-tests of MLAT) and the criterion measure Σ(MLA). The Ss will be classified as in the following 2 X 3 Table and "Validity Coefficients" will be computed within each subgroup - (Achievement Motivation and Integrative Orientation measures simply used as classification devices).

Table 1

Classification of Ss According to Three Achievement Motivation Levels (High, Middle, and Low 1/3's) and Two Integrative Orientation Levels (High and Low).

		Achievement Motivation			Total
		Low 1/3 Group	Middle 1/3 Group	High 1/3 Group	
Integrative Orientation	High Group	A	B	C	G
	Low Group	D	E	F	H
	Total	I	J	K	T

Note: For the following Problems refer to Table 1.

Problem 3. Compare validity coefficients at two levels of Integrative Orientation (G to H).

Problem 4. Compare validity coefficients at three levels of Achievement Motivation (I to J to K).

Problem 5. Compare validity coefficients at various combinations of Integrative Orientation and Achievement Motivation levels. (A to B to C to D to E to F).

Problem 6. Compare validity coefficients within sub-groups A, B, C, D, E, F, G, H, I, J, and K to the validity coefficient computed on the total group (T).

Four other problems will also be investigated.

Problem 7. To test whether using Integrative Orientation as a "moderator variable" has any advantage or disadvantage over using this variable as a "predictor variable" in a multiple regression equation; compare the latter (R: criterion vs aptitude predictors, and orientation in total group) to R's criterion vs aptitude predictors (sub-groups G and H separately).

Problem 8. To test whether using Achievement Motivation as a "moderator variable" has any advantage or disadvantage over using this variable as a "predictor variable" in a multiple regression equation; compare the latter (R: criterion vs aptitude predictors, and motivation in total group) to R's criterion vs aptitude predictors (sub-groups I, J, and K separately).

Problem 9. To test whether using both Integrative Orientation and Achievement Motivation as "moderator variables" has any advantage or disadvantage over using these variables as "predictor variables" in a multiple regression equation; compare the latter (R: criterion vs aptitude predictors, orientation and motivation in total group) to R's: criterion vs aptitude predictors, (in sub-groups A, B, C, D, E, and F separately).

Problem 10. To provide a test of Vroom's[84] multiplicative or interactive model the expectation would be that attitudes which are measures of Ss motivational level, as Integrative Orientation and Achievement Motivation appear to be, will be more highly related to performance for high than for low ability Ss.

These 10 problems are re-stated in terms of the null hypotheses as follows:

Null Hypothesis:

1. There is no significant difference in the cross-validated Multiple R's comparing a Full Multiple Regression model to a Stepwise Multiple Regression model.

2. There is no significant improvement in prediction adding measures of Integrative Orientation and Achievement Motivation to a standard aptitude battery in a multiple regression equation.

3. There is no significant difference in prediction for Ss classified according to levels of High and Low Integrative Orientation.

---

84 Vroom, Op. Cit., p. 331.

Null Hypothesis: (Cont'd)

4. There is no significant difference in prediction for Ss classified according to Low 1/3, Middle 1/3 and High 1/3 levels of Achievement Motivation.

5. There is no significant difference in prediction for Ss classified according to various combinations of Integrative Orientation and Achievement Motivation levels.

6. There is no significant difference in validity coefficients for Ss classified in sub-groups as in 5 above and the validity coefficients computed for the total group (aptitude predictors only).

7. There is no significant difference in prediction whether using Integrative Orientation as a "moderator variable" or as a "predictor variable" in a multiple regression equation.

8. There is no significant difference in prediction whether using Achievement Motivation as a "moderator variable" or as a "predictor variable" in a multiple regression equation.

9. There is no significant difference in prediction whether using Integrative Orientation and Achievement Motivation as "moderator variables" or as "predictor variables" in a multiple regression equation.

10. There is no significant difference in terms of performance between High and Low ability Ss classified into High/Low groups on Integrative Orientation and High 1/3, Middle 1/3 and Low 1/3 Groups on Achievement Motivation.

This research is primarily directed to test Guion's[85] suggestion that measures of motivation in particular might serve better as moderators than predictors in selection research models. Null hypotheses 1 to 5, therefore, although relevant for certain decisions prior to testing null hypotheses 6 to 9 inclusive, are not to be considered as the main object of the research; their functions are analogous to that of lemmas in geometry.

As a second aim the research was also designed to test Vroom's[86] interactive or multiplicative model, i.e., Performance (P) = Ability (A) × Motivation (M), thus emphasizing the relevance of null hypothesis 10.

---

85 Guion and Gottier, Op. Cit., p. 158.

86 Vroom, Op. Cit., p. 331.

C H A P T E R 2

EXPERIMENTAL DESIGN

This chapter presents the methodology of the research. It begins with descriptions of: the subjects participating in the research, the measuring instruments used, and the context of the Canadian Public Service French Language Training Program. Following this is an outline of the testing, scoring, and data processing procedures. The methods used to obtain reliability estimates for all variables are presented, and the chapter concludes with the statistical procedures involved in testing the hypotheses enumerated at the termination of Chapter 1.

1. The Subjects

The Ss were 260 males and 40 females enrolled at Level 1 of Phase 1 of an intensive French Language Training Program in the Public Service Language Bureau schools of Ottawa, during the period January to July 1970. The age range of Ss was 23-61. A frequency distribution by age is shown in Appendix 1.

In terms of education, 45% of Ss possessed an academic level of high school graduation or less while 55% were university graduates. Time since leaving school ranged from two years to over two decades.

Because of the diversity of job descriptions in the Public Service, no attempt was made to classify Ss on the basis of job category. As a general description, the majority of Ss were at the lower and middle management levels with a salary range of approximately $9,000.00 to $20,000.00 per year.

Data were not available on the geographical distribution of Ss but it would be expected that most Canadian provinces would be adequately represented. As a summary description, it would appear that the sample is heterogeneous with regard to age, education, and job classification. All Ss were assumed to be fluent in English and although the majority had French courses in some phase of their earlier academic programs, their level of proficiency in French was considered to be extremely low. This is further substantiated by the fact that all Ss were assigned to the elementary phase of the first level of intensive French language training. This elementary phase consisted of an intensive audio-lingual course of eight hours per day for a period of three weeks (120 hours). The actual content of the course and the methods adopted for the training program will be described in greater detail in a subsequent section of this chapter.

2. The Measuring Instruments.

a. Criterion Measures

The MLA - Cooperative Foreign Language Achievement

Tests, French Version Form LB (MLA-LB)[1] were administered at the end of the third week of the intensive French Language Training to measure attained proficiency. Since in this research the subjects are commencing language learning the LB version was considered more appropriate than the more advanced MA and MB versions. In the authors' minds, the MLA -- Cooperative Foreign Language Tests measure:

> .... the learner's knowledge of structures that are common in speech and that are common in writing, and they measure his knowledge of vocabulary and idiom in their respective and productive aspects.[2]

The achievement test battery is divided into four parts as follows:

a. Listening Test: The subject (individually or in group) listens to taped material and then answers multiple-choice questions. Working time: 25 minutes;

b. Speaking Test: The subject receives his instructions from a master tape at an individual recording station. He responds by speaking into a recorder. The test presents both aural and visual stimuli. Working time: 10 minutes;

c. Reading Test: The subject answers multiple-choice questions. Sentence completion items are intended to measure the understanding of words and idiomatic expressions; reading passages to measure the ability to understand main ideas, to find details, and to draw conclusions. Working time: 35 minutes;

---

1 Manual MLA -- Cooperative Foreign Language Tests, French Form LB, Princeton: Educational Testing Service, Cooperative Test Division, 1969, 32p.

2 Idem, ibid., p. 20.

d. Writing Test: The subject demonstrates his writing ability by writing both structured and free responses. Working time: 35 minutes.

b. Predictor Measures

The standard aptitude measures were chosen to include tests which, in the review of the literature, were found to be most effective in predicting performance in intensive second language learning. These were the five sub-tests of Carroll-Sapon's[3] Modern Language Aptitude Test (MLAT).

The MLAT is an outcome of a five-year research study conducted during the years 1953-58 at Harvard University. The test consists of five relatively uncorrelated sub-tests which have consistently shown good validity and made unique contributions to the prediction of success in foreign language training. It has been found to be particularly useful in predicting success in learning to speak and understand a foreign language, but it is also useful in predicting success in learning to read, write, and translate a foreign language.

The complete test requires about sixty to seventy minutes to administer. A tape recorder is used to present the instructions and test stimuli. As nearly as can be

---

3 John B. Carroll and Stanley M. Sapon, Modern Language Aptitude Test -- MLAT, Manual 1959 Edition, the Psychological Corporation, New York, p. 3-27.

inferred from factor analysis[4] and from a knowledge of test content, the traits measured by the five part scores are described as follows:

(i) Part 1. Number Learning. Seems to measure one aspect of the memory component of foreign language aptitude -- but the part has a fairly large specific variance, which one might guess to be a special "auditory alertness" factor which would play a role in auditory comprehension of a foreign language.

(ii) Part 11. Phonetic Script. This appears to measure what has been called "Sound-Symbol Association Ability", i.e., ability to learn correspondences between speech sounds and orthographic symbols. It may also measure a sort of memory for speech sounds and it tends to correlate highly with the ability to mimic speech sounds and sound combinations in foreign languages;

(iii) Part 111. Spelling Clues. Scores on this part depend to some extent on the student's English vocabulary knowledge. This sub-test also measures the some kind of sound -- symbol association ability as measured by Part 11, Phonetic Script, but to a lesser extent. It is highly speeded.

(iv) Part 1V. Words in Sentences. This part is thought to measure sensitivity to grammatical structure, and may be expected to have particular relevance to the student's ability to handle the grammatical aspects of a foreign language.

(v) Part V. Paired Associates. This part measures the rote memory aspects of the learning of foreign languages.

---

4 J.B. Carroll, "A Factor Analysis of Two Foreign Language Aptitude Batteries", in Journal of General Psychology, Vol. 59, 1958, p. 3-19.

The standardized tests that purport to measure the achievement motive were chosen among those which in the review of the literature were found to be effective in predicting performance in intensive second language learning. These were the three scales of Gough's[5] California Psychological Inventory (CPI) grouped under the broad category of Measures of Achievement Potential and Intellectual Efficiency. Particularily relevant to the selection of the CPI scales was the research by Datel et al[6] at the United States Army Language School. The findings of this research have been discussed in detail in the review of the literature.

The three CPI scales are described as follows:

(i) Achievement Via Conformance (Ac): Subjects are asked to indicate whether they agree or disagree with 37 descriptions of personal opinion such as, "I have a strong desire to be a success in the world" (agree), "I am said to be hotheaded" (disagree). The statements were designed to identify "those factors of interest and motivation which facilitate achievement in any setting where conformance is a positive behavior." A high score means that the subject is "capable, co-operative, efficient, organized, responsible, stable and sincere, is persistent and industrious, valuing intellectual activity and intellectual achievement." A low score means that the subject

---

5 Harrison G. Gough, Manual for the California Psychological Inventory, Palo Alto, California: Consulting Psychologists Press, 1957.

6 William E. Datel, Forest D. Hall, and Charles P. Rufe, "Measurement of Achievement Motivation in Army Security Agency Foreign Language Candidates", in Educational and Psychological Measurement, Vol. XXV, No. 2, 1965, p. 539-545.

is "coarse, stubborn, aloof, awkward, insecure and opinionated, is easily disorganized under stress or pressures to conform and pessimistic about his occupational future."

(ii) Achievement Via Independence (Ai): Subjects are asked to indicate whether they agree or disagree with 31 descriptions of personal opinion, for example, "I look up to my father as an ideal man" (disagree); "I seem to be about as capable and smart as most others around me" (agree). The statements were designed to identify "those factors of interest and motivation which facilitate achievement in any setting where autonomy and independence are positive behaviors." A high score means that a subject is "mature, forceful, strong, dominant, demanding and forsighted, is independent and self-reliant and having superior intellectual ability and judgment." A low score means that the subject is "inhibited, anxious, cautious, dissatisfied, dull and wary, is submissive and compliant before authority and lacking in self-insight and self-understanding."

(iii) Intellectual Efficiency (Ie): Subjects are asked to indicate whether they agree or disagree with 48 descriptions of personal opinion, for example, "I gossip a little at times" (agree); "people often expect too much of me" (disagree). The statements were designed to indicate the degree of personal and intellectual efficiency which the individual has attained. A high score means that the subject is efficient, clear-thinking, intelligent, progressive, thorough and resourceful; alert and well-informed; placing a high value on intellectual matters. A low score means that the subject is confused, cautious, easygoing, defensive, shallow and unambitious; conventional and sterotyped in thinking, lacking in self-direction and self-discipline.

The standard attitudinal or orientation measure selected for the research includes a scale developed by

Lambert[7] to measure Ss' integrative reasons for studying French. Subjects are asked to rate on seven point scales the extent to which each of four integrative reasons for studying French is descriptive of their own feelings. An integrative reason is one which places emphasis on learning a second language in order to learn more about, or to meet members of the other language group. For example: "It will enable me to gain good friends more easily among French speaking people". A high score indicates that integrative reasons are characteristic of the individual, a low score that they are not. A copy of the Integrative Orientation Measure is reported in Appendix 3.

3. The Context of the French Language Training Program.

Regarding the material to be selected for a French language training program, Carroll made the following recommendation:

> Useful experiments in foreign language teaching can be conducted by adhering fairly closely, at least initially, to patterns of teaching and types of teaching material which have already been developed and found successful by foreign language teachers. Foreign language teachers sincerely want to know how to improve their procedures, but they are not likely to attend carefully to the results of researchers which utilize a radically different

---

7 Robert C. Gardiner and Wallace E. Lambert, "Motivational Variables in Second-Language Acquisition", in Canadian Journal of Psychology, 1959, 13 (A), p. 267.

> or unrealistic teaching methods, or which appear to be formulated on terms of outmoded conceptions of language or language teaching.[8]

Voix et images de France (VIF) would seem generally to fulfill Carroll's[9] recommendation. The course has widespread use and recognition, and has been taught at the Public Service Language Bureau School[10] for five years.

On the subject and material for such a program there is some reservation concerning the relevance of VIF to the Language Bureau needs. This reservation is outlined by Blackburn as follows:

> The first and most pressing concern was the selection of suitable curriculum material. There are many programs on the market and most of these serve the purposes for which they were designed. A start was made in 1964 using a melange of what was thought to be the best of these systems. It became clear, however, that it would be necessary to construct a pedagogical and linguistic program specially to meet the needs of literate adults in a Canadian bicultural environment with particular attention being given to the social and administrative needs of public servants.[11]

The method of teaching VIF has been organized to

---

8 John B. Carroll, Research on Teaching Foreign Languages, Handbook of Research on Teaching, N.L. Gage, Ed., New York: Rand McNally and Company, 1963, p. 1065-1066.

9 Idem, ibid., p. 1065-1066.

10 G.A. Blackburn, A Bilingual and Bicultural Public Service, Canadian Public Administration Order 12, 1969, p. 5.

11 Idem, ibid., p. 5.

include four phases[12] as follows: "Présentation", "Explication", "Répétition", and "Transposition". In adopting the method to its needs, the Language Bureau blended in two other phases: "Exploitation", and "Fixation".

The first phase "Présentation" is explained by Renard and Heinle as follows:

> The purpose of the presentation is to provide students with a "global understanding" (general comprehension) of the situation. The dialogue will be interpreted (understood) in terms of the students' previous acquisition and their perception of the non-verbal components of the situation.[13]

The picture illustrates the sound segment; it becomes the link between the situation and the expression. Renard[14] suggests that the picture should appear slightly before the sound segment is heard and remain on the screen slightly after the sound segment is stated.

In the next phase, "Explication", the teacher verifies comprehension and explains what is still unknown. Drawings of situations are used to convey meaning and translation is avoided. The teacher proceeds from the known to the unknown, from the concrete to the abstract, from the easy to the

---

12 Collette Renard, and Henry Charles Heinle, Implementing Voix et Images de France in American Schools and Colleges Part 1., Philadelphia: Chilton Books, 1969.

13 Idem, ibid., p. 35.

14 Idem, ibid., p. 35.

difficult. He also limits himself to the students' already acquired vocabulary and structures. The goal for the learner is a coordinate system described by Brooks as follows:

> "... a coordinate system of two languages in which not only the overt patterns of behavior that characterize the new language, but also the mental processes that accompany it, shall have equal status with the mother tongue, yet be entirely separate from it"[15]

In the first step of the next phase, "Répétition", the teacher plays the tape and the students repeat. Then, without use of the tape, the student reconstructs the dialogue using the illustrations. Once the dialogue is thoroughly memorized, the teacher gives each student a role. When an illustration is flashed on the screen each student plays his role in the most realistic manner possible.

The fourth phase, "Exploitation avec images", involves the systemization of the new material, in order to create skills. The new material here means the important structures in the last lesson, as well as structures in previous lessons that have presented problems. Following is an example from the Répertoire de la méthodologie of "Exploitation avec images".

> Prof. : (En montrant les maisons) Qu'est-ce que c'est?
> Elève : Ce sont des maisons.
> Prof. : (En montrant la place) C'est une rue?
> Elève : Non, c'est une place.
> Prof. : (En montrant Pierre) Voilà Jacques?
> Elève : Non, voilà Pierre.

---

15 Nelson Brooks, Language and Language Learning: Theory and Practice, Second Edition, New York: Harcourt Brace and Company 1964, p. 49.

The next phase is "Exploitation sans images". The structures and the elements remain the same but they are used in a freer context. It is a matter, then, of changing the situations rather than the structures. The technique to achieve "Fixation" is to use a stimulus that is either a micro-situation or a simple question. Teachers are provided with the structures which are part of that lesson bloc. Then questions are developed in such a way that structures will be presented in no particular order.

The sixth and last phase is "Transposition". It relies on the premise that it is not sufficient that a student have automatic use of certain structures; he must, additionally, be able to use them in various contexts. To achieve this, the teacher creates the scene for the student. If it is a context in which the student's vocabulary may be wanting, the teacher supplies the word but he need not expect the student to have more than a passive meaning. The teacher is the stage director, the students, the actors. The teacher may want to create very simple mini-scenes for the less advanced students. At the beginning, "Transposition" can present a real problem to the student. The Handbook notes:

> Donc, après une parfaite systématisation, il ne faut pas s'attendre à ce que l'élève puisse facilement s'exprimer même si le cadre suggéré est admirablement trouvé pour remplir des structures qui ont été maîtrisées. Donc ces exercises seront d'abord très laborieux, mais après trois

ou quatre leçons on verra progressivement ces exercises porter fruits.[16]

Although the method is clearly structured it does not eliminate the possibility of a teacher's being creative and imaginative. But the creation and imagination are channelled within clear limits.

4. Procedures - Testing, Scoring and Data Gathering

The Modern Language Aptitude Test (MLAT) was administered to all Ss prior to their undergoing intensive French language training. In administering the MLAT[17], a tape recorder was used to present the instructions and test stimuli. Ss were seated in individual recording booths in the Public Service Language Laboratory, each subject having individual ear sets and direct contact with the master recording and relaying station. The presentation of the test via tape recorder ensured standard administration for all Ss participating in the research.

Achievement via Independence (Ai), Intellectual Efficiency (Ie), and Achievement via Conformance (Ac) are the three scales of Gough's[18] 1957 California Psychological

---

16 Idem, ibid., p. 13.

17 Carroll and Sapon, Op. Cit., p. 3.

18 Gough, Op. Cit., p. 3.

Inventory (CPI) which purport to measure academic potential and achievement motivation. The items comprising the scales, 105 in all, were chosen from the total pool of 480 items making up the CPI and were reproduced in test booklets with the usual "agree-disagree" instructions. The three scales so constituted were administered to all Ss who subsequently went on to study the French language in the Public Service Language Schools. This procedure was similar to the procedure adopted by Datel et al[19] in their U.S. Army Defense Language Institute Study.

A final predictor variable - a measure of Integrative Orientation, which was also used in the McGill[20] studies, was mimeographed on a 8 X 11 sheet with the usual instructions. This measure was also obtained prior to Ss undergoing intensive French language training.

After scoring all predictor tests, results were checked and punched on IBM Hollerith cards for subsequent computer processing. The test results were carefully guarded to ensure that they were not made available to administrative and training staffs in the language schools.

Subsequent to an intensive training program of three weeks, the Modern Language Achievement Tests - French Versions

---

19 Datel et al, Op. Cit., p. 539-545.

20 Gardiner and Lambert, Op. Cit., p. 267.

- LB (MLA-LB)[21] was administered to 300 Ss who completed phase 1 of level 1 of the language training program. Administration was conducted in the following manner:

1. For both the Reading and Writing sub-tests of the Achievement Battery, Ss were administered the test in groups of 10-30 in a standard classroom setting;

2. For the Listening sub-test of the MLA-LB, the S was seated in an individual recording booth in the Language Laboratory, listening to taped material and answered multiple-choice questions on a separate answer sheet; and

3. In the Speaking sub-test of the achievement battery, the S received his or her instructions from a master tape at an individual recording station and responded by speaking into a recorder. Both aural and visual stimuli were presented.

Since the responses to the Reading and the Listening MLA-LB sub-tests were recorded on separate answer sheets the scoring was conducted using hand-scoring stencils. The Speaking and Writing sub-tests required the services of trained linguists and these were provided by the Public Service Language Bureau from the Department of Linguistics at the University of Ottawa. Again, as in the case of the predictor variables, all results were carefully checked, coded and punched on IBM Hollerith cards for subsequent computer analysis.

---

21 MLA-LB, Op. Cit., 32p.

5. The Methods Used to Obtain Reliability Estimates.

Since an estimate of reliability is specific to the group on which it is obtained experimentally, it must always be interpreted in a relativistic manner. Because of such conditions as range of ability present in the population samples, type of operation used in determining reliability, etc., it can rarely be said that the reliability of a test is of a certain value. This caution is stated succinctly by Guilford as follows:

> One should speak of the reliability of a certain instrument applied to a certain population under certain conditions.[22]

No attempt, therefore, has been made in this research to document stated reliabilities from test manuals, research publications, etc; rather reliability was determined empirically on a sample of the available population.

The reliabilities of the MLA-LB criterion measures was estimated by computing an alternate form reliability estimate between MLA-LA administered to a sample of the population prior to their undergoing language training and MLA-LB administered to the same group three weeks later.

---

22 J.P. Guilford, Fundamental Statistics in Psychology and Education, Fourth Edition, McGraw-Hill Book Company, Toronto, 1965, p. 439.

Since alternate forms were not available for the predictor tests the reliability estimates were test-retest values based on a sample of approximately 40 Ss who returned for phase two of level one of the Public Service French language training program. The time between initial testing and retesting was nine weeks.

For all reliability estimates reported, means, SD's and SEM's were also computed. The latter statistics were presented to permit a more realistic interpretation by the reader of the reliability estimates obtained.

6. Statistical Analyses

Prior to elaborating the statistical operations used to test the null hypotheses outlined at the end of Chapter 1 it was considered necessary in this research to explore initially the relationships among the different sets of predictor and criterion variables. In the prediction context, there are practical advantages to having as few predictors as possible. The investigator is usually interested in maximizing the predictability of the criterion in subsequent samples. A subset of predictors may have better cross-validity than the use of all predictors. There is both theoretical and empirical justification for this expectation. In discussing limitations

in the multiple regression approach, Guilford[23] discusses three problems that should be considered: phenomenon of linear restraint, shrinkage, and assumption of linear regression.

It has been a common experience by users of multiple regression techniques that once three or four of the most valid tests have been combined in a multiple regression equation to predict a criterion, adding more tests rarely improves prediction. This is stated by Guilford[24] to be due to the phenomenon of linear restraint. One way to reduce the number of restraints according to Cureton[25] is to apply multiple regression techniques after removing as many restraints as possible by combining tests that intercorrelate highly with one another and have them enter the battery as one variable. Accordingly, in this research particular attention was directed to all variables (both predictor and criterion variables) to determine what tests or sub-tests could reasonably be combined.

As also stressed by Guilford[26], since multiple

---

23 J.P. Guilford, Psychometric Methods, Second Edition, McGraw-Hill Book Company, Inc., Toronto, 1954, p. 404.

24 Idem, ibid., p. 404.

25 E.E. Cureton, "Symposium: The Needs and Means of Cross-Validation", in Educational and Psychological Measurement, 1951, 11, p. 12-15.

26 Guilford, Op. Cit., 1954, p. 405.

regression procedures present an overoptimistic picture of predictive value of tests by capitalizing on chance, the best indication of shrinkage is actually to use the results on a new sample. This means cross-validation and as previously emphasized by McCornack[27] in the chapter on the review of the literature, it was accepted that no method is known that will select a subset of predictors that will be optimum in cross-validation. In this research, therefore, cross-validation was carried out where the sample size was considered large enough to warrant such an operation.

On the problem of the assumption of linear regression, Guilford[28] points out the actual and potential existence of many curvilinear regressions, in particular in the case of interest and temperament scores. At any rate he states that it behooves the investigator who employs multiple regression procedures to assure himself that his regressions are linear. For this research it was assumed that the regressions of achievement on aptitude and orientation were linear but for the measures of Achievement Motivation, correlation ratios in addition to Pearson correlation coefficients were computed and

---

27 Robert L. McCornack, "A Comparison of Three Predictor Selection Techniques in Multiple Regression" in Psychometrika, Vol. 35, No. 2, June 1970, p. 257-271.

28 Guilford, Op. Cit., 1954, p. 404.

compared. In the case of regression of aptitude on motivation and vice versa (since directionality not known) similar comparisons were made. In computing the correlation ratios and Pearson r's the steps as outlined in Dayhaw[29] were followed. The assumption of linearity of regression was tested by comparing the Pearson r's between the predictor and criterion variable (or other predictor) and the corresponding correlation ratio.

To test Null Hypothesis 1, that there is no significant difference in the cross-validated multiple R's comparing a Full Multiple Regression Model to a Stepwise Multiple Regression Model, a double cross-validation design was adopted. Using a table of random digits the total N of 300 was divided into two random samples: Sample A and Sample B consisting of 150 Ss each. For the full Multiple Regression Model all seven predictors were weighted simultaneously and a multiple R computed for Sample A. The regression weights obtained on Sample A were then applied to Sample B and a cross-validated multiple R obtained. Similarily a multiple R was computed for Sample B and the regression weights obtained applied to Sample A. The same procedure was adopted for the Stepwise Model and cross-validated multiple R's obtained for both

---

29 L.T. Dayhaw, Manuel de Statistique, 3e édition, Editions de l'Université d,Ottawa, Ottawa, Canada, 1966, p. 128-134.

Samples A and B[30]. To test the significance of all multiple R's, Guilford's table of significant correlation coefficients was used[31]. In order to test the significance of the difference between the cross-validated multiple R's for Full and Stepwise Multiple Regression Models the following formula from Guilford[32] was used.

$$F = \frac{(R^2 - R^2_2)\quad (N - M_1 - 1}{(1 - R^2_1)\quad (M_1 - M_2)}$$

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.

In the use of the F tables, the df_1 degrees of freedom are given by $(M_1 - M_2)$ and df_2 degrees of freedom by $(N - M_1 - 1)$.

To evaluate Null Hypothesis 2, that there is no significant improvement in prediction adding measures of Integrative Orientation and Achievement Motivation to the standard aptitude

---

30 Computer programs for both the Full Model and Stepwise Multiple Regression Model were supplied and processed at the University of Ottawa Computing Centre through the courtesy of Dr. Olgierd Porebski.

31 Guilford, Op. Cit., p. 580-581.

32 Idem, ibid., p. 402.

test battery in a multiple regression equation, a procedure suggested by Marks et al[33] was adopted. According to these researchers the only sure way of determining the extent to which a predictor will increase the predictive efficiency of a set of predictors is to compute the multiple R both with and without the predictor in question. For Null Hypothesis 2, therefore, multiple R's were computed for: aptitude, orientation and motivation predictors; aptitude and orientation predictors; aptitude and motivation predictors; and aptitude predictors only. The null hypothesis of no difference between the multiple R's was tested by the F test[34] since the multiple R with the larger number of variables included the same variables as the multiple R's with the smaller number of variables.

For Null Hypothesis 3, that there is no significant difference in prediction for Ss classified according to levels of High and Low Integrative Orientation, the total group was divided at the median on the Integrative Orientation measure and multiple R's computed independently between the aptitude predictors and the criterion measure. Since a multiple R has

---

33 M.R. Marks, R.E. Christal, and B.A. Bottenberg, "Simple Formula Aids for Understanding the Joint Action of Two Predictors", in Journal of Applied Psychology, 1961, 45, p. 287.

34 Guilford, Op. Cit., 1965, p. 402.

according to McNemar[35] the same meaning in a predictive sense as an ordinary bivariate correlation coefficient, both multiple R's were transformed to z's, and the standard error of the difference between the two z's was obtained in the usual manner[36].

To test Null Hypothesis 4, that there is no significant difference in prediction for Ss classified according to Low 1/3, Middle 1/3, and High 1/3 levels of Achievement Motivation, the total group was divided into three equal groups on the basis of the motivation measure. Three groups were chosen on this latter measure after it was found in the preliminary analysis that the regression of the achievement criterion on motivation was curvilinear. If Ss were classified into high and low groups on the basis of the motivation measure there would be the possibility that any improvement in prediction could be masked. The testing of Null Hypothesis 4 required the computation of multiple R's for each of the three groups. Again as for Null Hypothesis 3 the multiple R's were computed between the aptitude predictors and the criterion measure. Since in this case it was necessary to compare three multiple R's, all values were

---

35 Quinn McNemar, Psychological Statistics, Fourth Edition, 1969, John Wiley and Sons, Inc., Toronto, p. 194.

36 Idem, ibid., p. 157-158.

transformed to z's and the standard error of the difference was determined by reference to Edwards[37] using χ^2 test for homogeneity of k values of R.

For Null Hypothesis 5, that there is no significant difference in prediction for Ss classified according to various combinations of Integrative Orientation and Achievement Motivation levels, multiple R's were computed between the aptitude predictors and the MLA-LB criterion measure for the six groups categorized as A, B, C, D, E and F in Table 1 of Chapter 1. In this case, as for Null Hypothesis 4, the six multiple R's were transformed to z values and the standard error of the difference determined as in Edwards[38] using again the χ^2 test of homogeneity of k values of R.

To test Null Hypothesis 6 to 9 inclusive required the comparison of the validity coefficient of a subgroup to the validity coefficient of the total group. In a recent theoretical article, Zedeck[39] states that he does not know of a statistic that tests for the difference between the correlation

---

37 Allan L. Edwards, Experimental Design in Psychological Research, Revised Edition, Holt, Rinehart and Winston, New York, 1962, p. 83-84.

38 Idem, ibid., p. 83-84.

39 Sheldon Zedeck, "Problems with the Use of Moderator Variables", in Psychological Bulletin, 1971, Vol. 76, No. 4, p. 295-310.

coefficient in a subgroup and the correlation coefficient in the total group. He further criticizes research studies that reported conclusions based on such analyses, but that did not indicate the statistical tests that were applied. Zedeck further states:

> Logically, this type of test would be most desirable to determine whether moderators are adding to prediction. Yet, if we compare the two validity coefficients, the best we can do is "eyeball" the differences.[40]

As a tentative measure, Dayhaw[41] suggests the following formula by analogy with formulas dealing with part-whole differences of Means or Proportions[42].

$$\sigma_{Rs} = \frac{1 - R^2_t}{\sqrt{N_s - 1}} \sqrt{\frac{N_t - N_s}{N_t - 1}}$$

where

R_t = multiple R for total group;

N_t = number in total group;

N_s = number in subgroup.

This σ_{R_s} would be the standard error of the difference between

---

40 Idem, ibid., p. 303.

41 In a personal communication.

42 Lawrence T. Dayhaw, Manuel de Statistique, 3e édition, Editions de l'Université d'Ottawa, Ottawa, Canada, 1966, p. 318.

a multiple correlation coefficient of a subgroup and the corresponding multiple correlation coefficient in the total group which contains the subgroup. For Null Hypothesis 6, therefore, the above formula was applied to determine the standard error of the difference between the sub-group and total group multiple R's. For all comparisons (subgroups A-K inclusive, each to the total group, as in Table 1 of Chapter 1) the validity coefficients were computed between aptitude predictors and the MLA-LB criterion measure.

To test Null Hypothesis 7 the same procedure was adopted as for Null Hypothesis 6 except that only two comparisons (subgroups G, and H to total (T)) were made; also in this case Integrative Orientation was included in addition to the aptitude predictors in the computation of the validity coefficient for the total group.

To test Null Hypothesis 8 the same procedure was adopted as for Null Hypotheses 6 and 7 except that only three comparisons (subgroups I, J, and K to Total (T)) were made; in this case Achievement Motivation was included in addition to the aptitude predictors in the computation of the validity coefficient for the total group. The validity coefficient for total group was already computed in testing Null Hypothesis 2.

To test Null Hypothesis 9 that there is no significant difference in prediction between using Integrative Orientation and Achievement Motivation as "moderator variables" and using them as "predictor variables" in a multiple regression equation, the same procedure was adopted as for Null Hypothesis 7 and 8 except that in this case both the orientation and motivation predictors were included in computing the Multiple R for the total group. Again, referring to Table 1 of Chapter 1, multiple R's were computed between aptitude predictors and the MLA-LB achievement criterion for subgroups A to F inclusive and these validity coefficients were compared to the validity coefficients for the total group.

To test Null Hypothesis 10 that there is no significant difference in terms of performance for Ss classified according to: High and Low Ability (using total score on the MLAT), High/Middle/Low one-thirds on Achievement Motivation, and High/Low on Integrative Orientation, a 2 X 3 X 2 least squares analysis of variance fixed effects model design was utilized. This analysis resulted in 12 independent groups consisting of a minimum N of 21 and a maximum N of 29 in any one group. The performance scores, N's Means, SD's and variances for each group have been recorded in Appendix 4.

Although Box[43] has indicated the robustness of the F test with respect to violation of the assumption of homogeneity of variance under the usual conditions, in this study since the N's vary for the different subgroups and the bases for classification were subject variables, Hartley'[44] F Max Statistic was applied to the performance data in Appendix 4 to test for homogeneity of variance. The results of this analysis were considered prior to conducting an analysis of variance to test Null Hypothesis 10. Additionally, the performance means were plotted graphically for the different groups since it is recognized that statistical significance is largely a function of the size of N and with small N's, trends to significance should be at least observed.

7. Subsequent Analysis

In this research, it was found that the Full Multiple Regression with seven predictors had either superior or equal cross-validities to the Stepwise Model with three predictors.

---

43 G.E.P. Box, "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems", in the Annals of Mathematical Statistics, 25, p. 290-302.

44 Roger E. Kirk, Experimental Design: Procedures for Behavioral Sciences, Brooks/Cole Publishing company, Belmont, California, U.S.A., 1968, p. 62.

It would be expected, therefore, that when predictors are dropped and simply used to classify (moderators vice predictors) that the Full Model with five aptitude predictors would also cross-validate.

The cross-validation design requires a sufficient N and an inspection of the sub-groups referred to in Table 1 of Chapter 1 would indicate that for the motivation sub-groups (I, J, and K), and the orientation sub-groups (G and H) the numbers were sufficient to warrant cross-validation[45]. Since improved prediction was demonstrated within the Middle one-third Achievement Motivation group a double cross-validation was considered imperative for the subgroups designated as I, J, and K in Table 1. For each sub-group of 100 cases, using a table of random digits, two random samples of 50 Ss each were selected. For each sample, optimum weights were derived and a multiple R computed. The regression weights derived for one sample were then applied to the other sample. This resulted in six multiple R's and six cross-validated R's. Again the cross-validated R's were compared for the six samples using the χ^2 test for homogeneity of k values of R[46].

---

45 Jim C. Nunnally, Psychometric Theory, McGraw-Hill Book Company, Toronto 1967, p. 164-165.

46 Edwards, Op. Cit., p. 83-84.

Since for Vroom's[47] multiplicative model a significant interaction was not found (Null Hypothesis 10) and since the F values for: the two factor interactions -- Ability and Orientation (AC), Motivation and Orientation (BC); and the three factor interactions -- Ability, Orientation and Motivation (ABC) were significant only at the .25 level some pooling of groups was considered permissible[48]. The Orientation variable was dropped and the twelve groups re-categorized into six groups. A 2 X 3 least squares analysis of variance fixed effects was then conducted, and the necessary statistical tests applied.

In the following chapter, the results of the statistical treatments leading up to and testing the null hypotheses are presented and discussed. For rejection of the null hypotheses in this research a p level of .05 was adopted.

---

47 V.R. Vroom, Work and Motivation, John Wiley and Sons Ltd., New York, 1964, p. 331.

48 Kirk, Op. Cit., p. 214-215.

C H A P T E R III

PRESENTATION AND DISCUSSION OF RESULTS

This chapter, which presents and discusses the results of the analyses of the data, begins with a presentation and an interpretation of the reliability estimates for both criterion and predictor variables. For all reliability estimates - means, SD's and SEM's are also presented. Specific attention is given to the inter-relationships among the different sets of criterion and predictor variables to determine empirically what variables might reasonably be combined. In addition, tests of linearity of regression are made and presented for the motivation variable. Because of the controversy in the research literature over the utility of different regression models, the popularly used Stepwise Model is compared to a Full Multiple Regression Model in a double cross-validation design. Subsequently, the Full Multiple Regression Model is used to test null hypotheses designed to determine whether measures of motivation and orientation function better as moderators than as predictors in selection research models. A least squares analysis of variance is presented and discussed to test for interactive effects of ability, motivation and orientation on performance in intensive French language learning. Finally, after considering the results of the 2 X 3 X 2 analysis of variance, the orientation variable was dropped and a 2 X 3 analysis of variance computed

to test for interactive effects of ability and motivation.

1. Reliability

a. Reliability of the Criterion Measures:

The alternate form reliability (r_{af}), means, SD's and SEM's for the Modern Language Achievement Tests - French Version - LB (MLA-LB) are presented in Table 11. All of MLA-LB sub-tests except for the Speaking Test are observed to have a high degree of reliability. The total score reliability is also high. In terms of the classical and quite stringent guide set down by Kelley[1] for achievement tests many years ago, the reliabilities observed in Table 11 would generally meet the minimum requirements.

b. Reliabilities of the Predictor Measures:

Test re-test reliability coefficients, means, SD's and SEM's for aptitude tests, motivation, and orientation measures are presented in Table 111. All of the computed reliabilities except for the Number Learning sub-test of the Modern Language Test Battery and two sub-scales of the California Psychological Inventory-Achievement Independence (Ai) and Intellectual Efficiency (Ie) exceed .70. Although the reliabilities are not

---

1 T. Kelley, Interpretation of Educational Measurements, Yonkers, New York: World Book Company, 1927.

Table II

Means, SD's, Reliability Coefficients - Alternate Forms (r_{af}), SEM's for Criterion Variables (Time Between Testing - 3 weeks)

Statistics	MLA* (a) L	(b) S	(c) R	(d) W	(e) T
N	201	201	201	201	201
Means					
MLA-LA	23.11	39.25	28.92	31.58	129.86
MLA-LB	28.13	47.24	31.65	41.29	148.31
SD's					
MLA-LA	11.18	16.10	13.86	25.87	61.32
MLA-LB	9.07	12.66	12.08	29.75	58.03
r_{af}	.87	.70	.93	.91	.92
SEM's					
MLA-LA	4.63	8.82	3.57	8.20	17.32
MLA-LB	3.24	6.93	3.20	9.50	16.39

* MLA: Modern Language Achievement Cooperative Foreign Language Test - French Forms LA & LB

(a) L- Listening Test

(b) S- Speaking Test

(c) R- Reading Test

(d) W- Writing Test

(e) T- (Total Score)

Table III

Means, SD's, Test Re-Test Reliability Coefficients (r_{tt}), SEM's for Predictor Variables (Time Between Testing - 9 weeks)

STATISTICS	MLAT*						CPI#				Integrative Orientation
	(a) 1	(b) 11	(c) 111	(d) 1V	(e) V	(f) Total	(g) Ac	(h) Ai	(i) Ie	(j) Total	
N	40	40	40	40	40	40	41	41	41	41	41
Means											
1. Test	34.65	23.90	21.00	24.50	13.05	117.10	29.71	22.22	37.78	89.71	19.34
2. Re-test	36.95	25.63	27.08	25.83	17.05	132.45	29.44	21.59	36.76	87.78	19.29
SD's											
1. Test	7.76	3.34	7.93	6.57	5.87	21.60	4.24	3.67	4.53	10.14	4.14
2. Re-test	5.18	3.43	8.93	7.53	4.79	21.92	3.75	3.82	4.60	7.49	3.95
r_{tt}	.57	.71	.74	.87	.70	.81	.77	.68	.58	.74	.71
SEM's											
1. Test	5.09	1.80	4.00	2.70	3.30	9.79	2.50	2.20	2.94	5.00	2.20
2. Re-test	3.40	1.82	4.50	3.10	2.70	9.80	2.00	2.18	2.95	3.82	2.18

* MLAT: Carroll-Sapon's Modern Language Aptitude Test

(a) 1- Number Learning; (b) 11- Phonetic Script; (c) 111- Spelling Clues;
(d) IV- Words in Sentences; (e) - Paired Associates; (f) - Total: Summation Score

CPI: Gough's California Psychological Inventory (Achievement Motivation Measure)

(g) Ac: Achievement Conformity Scale; (h) Ai: Achievement Independence Scale;
(i) Ie: Intellectual Efficiency Scale; (j) - Total: Summation Score.

as satisfactory as those obtained for the criterion measures, they meet Kelley's[2] minimum requirements for the evaluation of level of group accomplishment. An evaluation of the data in Table III would suggest a restricted variability for some of the sub-tests, one factor that could affect the magnitude of the reliability estimates obtained. For this reason, the SEM's are also reported since such indices are independent of the variability of the group on which the reliability coefficient is computed[3]. It is acknowledged that the N's used to estimate the test re-test reliability estimates for the predictor variables were smaller than desired, but it was not possible in the time allotted in the training program to re-test more subjects.

2. Intercorrelations of Criterion and Predictor Variables

a. Intercorrelations of Criterion Variables:

Intercorrelations of the Listening (L), Speaking (S), Reading (R), and Writing (W) sub-tests of the MLA-LB Battery are shown in Table 1V. One critical consideration in an evaluation of the data is that if tests are highly intercorrelated there is statistical justification for combining them. The

---

2 Idem, ibid.

3 G.C. Helmstadter, Principles of Psychological Measurement, New York: Appleton-Century-Crafts, 1964, p. 76-77.

Table IV

Intercorrelations, Means, and SD's of the Listening (L), Speaking (S), Reading (R), and Writing (W) Sub-Tests of the MLA-LB Battery (N = 201)

	MLA-LB			
Variables	L	S	R	W
L	1.00	0.65	0.85	0.82
S	0.65	1.00	0.63	0.60
R	0.85	0.63	1.00	0.90
W	0.82	0.60	0.90	1.00
Means	28.13	47.24	31.65	41.29
SD's	9.07	12.66	12.08	29.75

Significance at $P < .01$ requires an $r \geq .23$

Significance at $P < .05$ requires an $r \geq .16$

data in Table IV are provided for the same groups on which the reliability data in Table II were based. This is the usual matrix, R, of product-moment correlations among the variables, with unity in the diagonal cells. It is often expressed in matrix algebra as:

$$R = \frac{1}{N} \; Z' \; Z$$

where Z is the matrix of standardized scores with subjects as rows and variables as columns and Z' is its transpose. The intercorrelations range from .60 to .90 which would indicate that in the early stage of language learning, performance in listening, speaking, reading and writing are highly correlated. This is consistent with Carroll's[4] findings in Peace Corps studies which prompts one to suggest that special abilities required for development of different skills in language learning may only manifest themselves later in the language training program. Since this research is restricted to the initial phases of language learning, the MLA-LB total score will be used as the criterion variable. This will also ensure the use of a highly reliable criterion measure (See Table II).

---

4 John B. Carroll, A Parametric Study of Language Training in the Peace Corps, Harvard University Press, Cambridge, Mass., 1966, p. 91.

b. Intercorrelations of Predictor Variables:

Again as for the criterion variable, a preliminary analysis was conducted to determine the intercorrelations of all predictor variables. The intercorrelations for the predictor variables are reported in Table V. An evaluation of the intercorrelations of the sub-tests of the Modern Language Aptitude Tests (MLAT), that is variables 1 to 5, would suggest that the separate parts do in fact as Carroll et al[5] claim, measure somewhat different aspects of foreign language aptitude. The intercorrelations are low enough to provide unique information. There is also Carroll's[6] finding that optimal weighting of the MLAT sub-tests in the multiple regression approach yields significantly better prediction of success in intensive language learning than the sum of raw scores. As a check on such a finding, a Pearson r of .45 (N =300) was obtained between total score on the MLAT and the MLA-LB criterion measure; for the same population the Full Multiple Regression Model (all sub-tests of the MLAT optimally weighted) yielded a multiple R of .54. Computing an F test

---

5 John B. Carroll, and Stanley M. Sapon, Modern Language Aptitude Test Manual, 1959 Edition, The Psychological Corporation, New York, p. 17.

6 Carroll, Op. Cit., 1966, p. 56.

Table V

Intercorrelations, Means, SD's of the Sub-tests of the MLAT, the Achievement Motivation Scales of the CPI, and Integrative Orientation (N = 58)

	Variables	1	2	3	4	5	6	7	8	9
1.	MLAT-Number Learning	1.00	0.36	0.39	0.42	0.39	-0.02	0.10	0.06	-0.05
2.	MLAT-Phonetic Script	0.36	1.00	0.37	0.16	0.12	-0.21	-0.08	-0.08	-0.03
3.	MLAT-Spelling Clues	0.39	0.37	1.00	0.27	0.30	-0.09	-0.01	-0.12	-0.02
4.	MLAT-Words in Sentences	0.42	0.16	0.27	1.00	0.32	0.16	0.41	0.26	0.03
5.	MLAT-Paired Associates	0.39	0.12	0.30	0.32	1.00	0.08	0.26	0.12	-0.09
6.	CPI - Ac	-0.02	-0.21	-0.09	0.16	0.08	1.00	0.51	0.73	0.08
7.	CPI - Ai	0.10	-0.08	-0.01	0.41	0.26	0.51	1.00	0.63	-0.16
8.	CPI - Ie	0.06	-0.08	-0.12	0.26	0.12	0.73	0.63	1.00	0.17
9.	Integrative Orientation	-0.05	-0.03	-0.02	0.03	-0.09	0.08	-0.16	0.17	1.00
	Means	28.6	21.6	20.4	21.5	12.4	28.9	21.3	37.6	17.5
	SD's	10.3	3.95	8.17	7.71	5.24	4.80	4.15	5.72	4.58

Significance at $P < .01$ requires an $r \geq .30$

Significance at $P < .05$ requires an $r \geq .21$

for this difference[7] yielded a value of 10.43 which is statistically significant for $p < .01$. Based on these data the sub-tests of the MLAT will be utilized as separate predictor variables in this research.

Variables 6,7, and 8 in Table V are three scales of the California Psychological Inventory: Achievement Conformity (Ac), Achievement Independence (Ai), and Intellectual Efficiency (Ie), that purport to measure the achievement motive[8]. An evaluation of the data in Table V would indicate a high degree of intercorrelation for the scales - Ac & Ai .51; Ac & Ie .73; Ai and Ie .63. According to Anastasi[9] this lack of independence, resulting in redundancy among the scores, is perhaps the chief limitation of the CPI. Further Datel et al[10] found that Ss in intensive foreign language training who scored high on the Ac also scored high on both the Ai and Ie

---

7 J.P. Guilford, Fundamental Statistics in Psychology and Education, Fourth Edition, McGraw-Hill Book Company, Toronto, 1965, p. 403.

8 H.G. Gough, Manual for the California Psychological Inventory, Palo Alto, California: Consulting Psychologists Press, 1957, 22p.

9 Anne Anastasi, Psychological Testing, Third Edition, The McMillan Company, Toronto, Ontario, 1968, p. 448.

10 William E. Datel, Forrest D. Hall, and Charles P. Rufe, "Measurement of Achievement Motivation in Army Security Agency Foreign Language Candidates" in Educational and Psychological Measurement, Vol. XXV, No. 2, 1965, p. 539-545.

scales. Accordingly, in this research scores obtained on the three CPI scales were summed and treated in the different selection models as one variable. Again, referring to Table III, the use of CPI Achievement Motivation total score also provides a more reliable measure (test re-test) than either the Ai or Ie scales used singly. (For raw score data see App. 2).

3. Linearity of Regression

Linearity of regression was tested by comparing the Pearson r and the eta for the regression of the MLA-LB criterion measure on the CPI Achievement Motivation measure. In the case for the regression of aptitude on motivation and vice versa (since directionality not known) similar comparisons were made. In computing the Pearson r's and the eta's the procedure as outlined in Dayhaw[11] were followed. The results of these analyses are shown in Tables VI and VII. As stated in the Chapter on Experimental Design it was assumed that the regressions of achievement on aptitude and achievement on orientation were linear.

From the data in Table VI and VII only one regression, that of the MLA-LB criterion measure on the CPI, is statistically significant ($P < .05$). The regression of the MLAT

---

11 L.T. Dayhaw, Manuel de Statistique, 3e édition, Editions de l'Université d'Ottawa, Ottawa, Canada, 1966, p. 130.

Table VI

Linearity of Regression of the MLA-LB Criterion Measure on the CPI Achievement Motivation Measure (N = 300)

Test	r	F_r	b_{yx}	b_{xy}	η_{yx}	$F_{\eta yx}$	$F(Lin)_{yx}$
CPI	.15	$7.31^{(1)**}$	.875	.025	.30	$2.33^{(2)**}$	$1.83^{(3)*}$

** (1) Significance at $P < .01$ requires an F value of 6.64

** (2) Significance at $P < .01$ requires an F value of 2.18

* (3) Significance at $P < .05$ requires an F value of 1.75

Table VII

Linearity of Regression of the MLAT Predictor Measure (y) on the Achievement Motivation Measure (x) and Vice Versa (N = 300)

Test	r	F_r	b_{yx}	b_{xy}	η_{yx}	$F_{\eta yx}$	$F(Lin)_{yx}$	η_{xy}	$F_{\eta xy}$	$F(Lin)_{xy}$
CPI	.12	4.64(1)*	.130	.118	.28	1.94(2)*	1.70	.058	.01	.21

*(1) Significance at $P < .05$ requires an F value of 3.84

*(2) Significance at $P < .05$ requires an F value of 1.75

predictor measure on the CPI although not statistically significant approaches significance. All of the regressions, therefore, except for the regression of the MLA-LB criterion measure on the CPI, are assumed to be linear.

4. Statistical Testing of Null Hypotheses 1-10 inclusive

Tables VIII and IX shows the intercorrelations, means, SD's and multiple R's for both validation and cross-validation samples for the Full Multiple Regression Model. The data in Tables VIII and IX would generally support the fact that validation Samples A and B are roughly equivalent.

Prior to conducting the double cross-validation, the first step was to compute multiple R's independently for both validation samples A and B. For the Full Multiple Regression Model the validity coefficients obtained for samples A and B were as previously reported .607 and .563 respectively (Tables VIII and IX). The regression weights obtained for Sample A were then applied to the scores for Sample B and conversely the weights obtained for Sample B applied to Sample A[12]. For the Stepwise Multiple Regression Model the same procedure was repeated and multiple R's and cross-validated R's derived for

---

12 Quinn McNemar, Psychological Statistics, Fourth Edition, John Wiley and Sons, Inc., Toronto, 1969, p. 208.

Table VIII

Intercorrelation Matrix, Means, SD's, and Multiple R for Predictor Variables (Sub-tests of MLAT, Integrative Orientation, CPI) and Criterion Measure (MLA-LB) for Validation Sample A (Full Multiple Regression Model N = 150)

Variables*	1	2	3	4	5	6	7	8
1	1.00	0.43	0.22	0.30	0.37	-0.03	0.04	0.30
2	0.43	1.00	0.34	0.38	0.25	-0.01	0.03	0.49
3	0.22	0.34	1.00	0.33	0.29	0.01	0.30	0.42
4	0.30	0.38	0.33	1.00	0.37	-0.11	0.22	0.41
5	0.37	0.25	0.29	0.37	1.00	-0.00	0.13	0.27
6	-0.03	-0.01	0.01	-0.11	-0.00	1.00	-0.07	0.07
7	0.04	0.03	0.30	0.22	0.13	-0.07	1.00	0.24
8	0.30	0.49	0.42	0.41	0.27	0.07	0.24	1.00
Means	31.13	22.42	19.12	21.47	12.87	19.36	86.56	147.75
SD's	10.25	4.11	8.96	7.72	5.57	4.11	8.95	54.38
Multiple R	0.607							

* Variables 1 to 5 inclusive: Sub-tests of MLAT
Variable 6 : Integrative Orientation
Variable 7 : CPI
} Predictors
Variable 8 : MLA-LB (Criterion)

Table IX

Intercorrelation Matrix, Means, SD's, and Multiple R for Predictor Variables (Sub-tests of MLAT, Integrative Orientation, CPI) and Criterion Measure (MLA-LB) for Validation Sample B (Full Multiple Regression Model N = 150)

Variables*	1	2	3	4	5	6	7	8
1	1.00	0.23	0.37	0.31	0.43	-0.02	-0.01	0.28
2	0.23	1.00	0.55	0.27	0.28	0.08	0.11	0.25
3	0.37	0.55	1.00	0.35	0.45	0.06	0.07	0.25
4	0.31	0.27	0.35	1.00	0.34	-0.02	0.12	0.53
5	0.43	0.28	0.45	0.34	1.00	0.09	-0.02	0.28
6	-0.02	0.08	0.06	-0.02	0.09	1.00	0.12	-0.02
7	-0.01	0.11	0.07	0.12	-0.02	0.12	1.00	0.12
8	0.28	0.25	0.25	0.53	0.28	-0.02	0.12	1.00
Means	30.11	22.53	19.77	22.32	12.73	18.97	87.68	142.12
SD's	10.08	4.04	9.06	7.13	5.89	4.35	10.83	60.62
Multiple R	0.563							

* Variables 1 to 5 inclusive: Sub-tests of MLAT
Variable 6 : Integrative Orientation
Variable 7 : CPI
} Predictors
Variable 8 : MLA-LB (Criterion)

both samples. These results are summarized in Table X.

The significance of all multiple R's and cross-validation R's in Table X was established by reference to Guilford[13]. All R's were found to be statistically significant at $P < .01$.

If the difference between the multiple R and the cross-validation R is defined as a "drop", it would seem legitimate to assume that the better model (i.e.- Full or Stepwise Model) would be the one that produces the smaller "drop" on cross-validation. From Table X it would appear that the Full and Stepwise models are equivalent when cross-validation is conducted from Sample A - to Sample B (.409 and .410 respectively). When cross-validation is conducted from Sample B - to - Sample A, however, the two regression models do not appear to be equivalent (.476 and .336). Since the multiple R's for the Full Multiple Regression Model and for the Stepwise Multiple Regression Model are about equal for Sample B (.563 and .556 respectively), testing the significance of the difference between the cross-validation R's (.476 vs .336) would give an indication of the difference in the "drop". In testing the significance of the difference between the cross-validated R's for the Full and Stepwise models, therefore, for Sample B, a

---

13 J.P. Guilford, "Fundamental Statistics in Psychology and Education", Fourth Edition, New York, McGraw-Hill, 1965, p. 580-581.

Table X

Multiple R's and Cross-Validation R's for Full Multiple Regression Model (Seven Predictors) and Stepwise Multiple Regression Model (Three Predictors) for Samples A and B.

Sample	Regression Model	Multiple R (N = 150)	Cross-Validation R (N = 150)
A	Full	.607**	A - to - B .409**
A	Stepwise	.583##	A - to - B .410##
B	Full	.563**	B - to - A .476**
B	Stepwise	.556##	B - to - A .336##

** Significance at $P < .01$ requires an $R \geq .324$

Significance at $P < .01$ requires an $R \geq .244$

table of the distribution of F was used[14]. The formula for computing F for this purpose reads:

$$F = \frac{(R^2_1 - R^2_2)\ (N - M_1 - 1)}{(1 - R^2_1)\quad (M_1 - M_2)}$$

where

R_1 =Multiple R with larger number of independent variable.

R_2 =Multiple R with one or more variables omitted.

M_1 =Larger number of independent variables; here 7.

M_2 =Smaller number of independent variables; here 3.

The F value obtained for the difference between the cross-validated R's (The Full and Stepwise models) for Sample B was 5.218. With 4 and 142 df an F value of approximately 3.44 is significant at $p < .01$. It was concluded from the data that the difference between the cross-validated multiple R's for the Full Model and Stepwise Model was statistically significant -- going from Sample B to Sample A; the Full Model cross-validated better than the Stepwise Model for Sample B. No statistically significant difference was observed for Sample A. The null hypothesis of no significant difference in cross-validated multiple R's between Full and Stepwise Multiple Regression Models is therefore accepted for Sample A, but rejected for Sample B.

For the next series of problems and/or hypotheses the

---

14 Idem, ibid., p. 586.

Full Model Multiple Regression (or restricted variations of the Full Model) will be adopted vice the Stepwise Model since it has been found that the Full Model cross-validated significantly better in one case (cross-validation B - to - A) than the Stepwise Model and was comparable in the second case (cross-validation A - to - B).

Table XI presents the intercorrelations, means, and SD's for all predictors: MLAT sub-tests, Achievement Motivation, Integrative Orientation; and the MLA-LB criterion measure. Full Model Multiple R's are given: a) for complete data; b) for Achievement Motivation data only omitted; c) for Integrative Orientation only omitted; and d) for both Achievement Motivation and Integrative Orientation omitted. This procedure according to Marks et al[15] is the only sure way of determining the extent to which a predictor will increase the predictive efficiency of a set of predictors.

The relevant data required for testing Null Hypothesis 2, that there is no significant improvement in prediction adding measures of Integrative Orientation and Achievement Motivation to the standard aptitude test battery (sub-tests of MLAT) in a multiple regression equation, is presented in the bottom section

---

15 M.R. Marks, R.E. Christal, and B.A. Bottenberg, "Simple Formula Aids for Understanding the Joint Action of Two Predictors", in Journal of Applied Psychology, 1961, 45, p. 287.

Table XI

Intercorrelation Matrix, Means, SD's and Multiple R's for: a) Complete Data, b) Variable 6 Only Omitted, c) Variable 7 Only Omitted, and d) Both Variables 6 and 7 Omitted, for Total Sample (N=300)

Variables*	1	2	3	4	5	6	7	8
1	1.00	0.33	0.29	0.30	0.40	0.02	0.01	0.30
2	0.33	1.00	0.44	0.33	0.27	0.04	0.07	0.36
3	0.29	0.44	1.00	0.34	0.37	0.04	0.17	0.33
4	0.30	0.33	0.34	1.00	0.36	0.07	0.17	0.47
5	0.40	0.27	0.37	0.36	1.00	0.05	0.05	0.29
6	0.02	0.04	0.04	0.07	0.05	1.00	0.03	0.03
7	0.01	0.07	0.17	0.17	0.05	0.03	1.00	0.17
8	0.30	0.36	0.33	0.47	0.29	0.03	0.17	1.00
Means	30.62	22.47	19.45	21.90	12.78	19.17	87.12	145.27
SD's	10.16	4.07	9.00	7.43	5.76	4.23	9.93	57.33

a) Multiple R using all 7 predictors 0.547
b) Multiple R omitting predictor 6 only 0.546
c) Multiple R omitting predictor 7 only 0.540
d) Multiple R omitting predictors 6 and 7 0.539

* Variables 1 to 5 inclusive: Sub-tests of MLAT
Variable 6 : Integrative Orientation } Predictors
Variable 7 : CPI
Variable 8 : MLA-LB (Criterion)

of table XI. In determining whether the multiple R with more independent variables included is significantly greater than the R's with smaller numbers of variables the F test[16] was used. In using the F table[17], the df_1 degrees of freedom are given by $(M_1 - M_2)$ and the df_2 degrees of freedom by $(N - M_1 - 1)$. None of the F values was found to be statistically significant. It was concluded from the data that there was no significant difference in multiple R's between the Full Model (all predictors) and the following restricted models: MLAT sub-tests and Integrative Orientation predictors, MLAT sub-tests and Achievement Motivation predictors, and MLAT sub-test predictors only. In other words the addition of the Integrative Orientation and Achievement Motivation predictors to the MLAT predictors did not significantly improve prediction. Null Hypothesis 2 is therefore accepted.

For Null Hypothesis 3, that there is no significant difference in prediction for Ss classified according to High and Low Integrative Orientation, multiple R's were computed independently for each group. The intercorrelation matrices, means, SD's, and multiple R's for the two groups are shown in Tables XII and XIII respectively. Both multiple R's were

---

16 Guilford, Op. Cit., 1965, p. 402.

17 Idem, ibid., p. 586.

Table XII

Intercorrelation Matrix, Means, SD's, and Multiple R for Predictor Variables (sub-tests of MLAT) for Ss Classified According to High Integrative Orientation (N = 150)

Variables*	1	2	3	4	5	6
1	1.00	0.38	0.30	0.29	0.44	0.31
2	0.38	1.00	0.48	0.40	0.34	0.41
3	0.30	0.48	1.00	0.33	0.39	0.31
4	0.29	0.40	0.33	1.00	0.39	0.50
5	0.44	0.34	0.39	0.39	1.00	0.37
6	0.31	0.41	0.31	0.50	0.37	1.00
Means	30.56	22.30	18.95	20.95	12.84	143.65
SD's	9.97	3.92	9.04	7.60	5.66	56.60
Multiple R	0.570					

* Variables 1 to 5 inclusive: Sub-tests of MLAT (Predictors)
Variable 6 : MLA-LB (Criterion)

Table XIII

Intercorrelation Matrix, Means, SD's and Multiple R, for Predictor Variables (sub-tests of MLAT) for Ss Classified According to Low Integrative Orientation (N = 150)

Variables*	1	2	3	4	5	6
1	1.00	0.29	0.29	0.31	0.36	0.28
2	0.29	1.00	0.41	0.25	0.20	0.32
3	0.29	0.41	1.00	0.35	0.35	0.34
4	0.31	0.25	0.35	1.00	0.32	0.44
5	0.36	0.20	0.35	0.32	1.00	0.21
6	0.28	0.32	0.34	0.44	0.21	1.00
Means	30.68	22.65	19.95	22.85	12.77	146.89
SD's	10.38	4.22	8.96	7.15	5.80	58.20
Multiple R	0.516					

* Variables 1 to 5 inclusive: Sub-tests of MLAT (Predictors)
Variable 6 : MLA-LB (Criterion)

transformed to Z's and the standard error of the difference between the two Z's was obtained[18]. For the data in Tables XII and XIII the Z coefficients corresponding to multiple R's of .570 and .516 are .648 and .577 respectively. The SE of either is .0825; the SE of their difference is .1166. The difference in the Z's, i.e., .071, was found to deviate from a difference of 0.00 to the extent of .071/.1166 = .609. For a two-tailed test with $P < .05$ this deviation is not significant. It is concluded, therefore, that the multiple R's for the High and Low Integrative Orientation groups do not differ significantly. Null Hypothesis 3 is therefore accepted.

To test Null Hypothesis 4, that there is no significant difference in prediction for Ss classified according to Low 1/3, Middle 1/3, and High 1/3 levels of Achievement Motivation, multiple R's were computed independently for each of the three groups. The intercorrelation matrices, means, SD's, and multiple R's for the three groups are shown in Tables XIV, XV and XVI respectively. In comparing the multiple R's, all values were transformed to Z's and the standard error of the difference was determined by reference to Edwards[19] using the χ^2 test for homogeneity of k values of R. The calculations for a

---

18 McNemar, Op. Cit., 1969, p. 157-158.

19 Allen L. Edwards, Experimental Design in Psychological Research, Revised Edition, Holt, Rinehart and Winston, New York, 1962, p. 83-84.

Table XIV

Intercorrelation Matrix, Means, SD's, and Multiple R for Predictor Variables (sub-tests of MLAT) for Ss classified According to Low 1/3 on Achievement Motivation (N = 100)

Variables*	1	2	3	4	5	6
1	1.00	0.32	0.21	0.35	0.44	0.19
2	0.32	1.00	0.42	0.36	0.17	0.40
3	0.21	0.42	1.00	0.32	0.29	0.23
4	0.35	0.36	0.32	1.00	0.40	0.39
5	0.44	0.17	0.29	0.40	1.00	0.17
6	0.19	0.40	0.23	0.39	0.17	1.00
Means	30.89	22.50	17.72	20.40	12.53	130.01
SD's	9.83	4.19	8.84	7.89	5.10	55.28
Multiple R	0.482					

* Variables 1 to 5 inclusive: Sub-tests of MLAT (Predictors)
Variable 6 : MLA-LB (Criterion)

Table XV

Intercorrelation Matrix, Means, SD's, and Multiple R for Predictor Variables (sub-tests of MLAT) for Ss Classified According to Middle 1/3 on Achievement Motivation (N=100)

Variables*	1	2	3	4	5	6
1	1.00	0.42	0.28	0.40	0.41	0.43
2	0.42	1.00	0.46	0.37	0.33	0.36
3	0.28	0.46	1.00	0.38	0.44	0.46
4	0.40	0.37	0.38	1.00	0.42	0.53
5	0.41	0.33	0.44	0.42	1.00	0.39
6	0.43	0.36	0.46	0.53	0.39	1.00
Means	30.33	22.03	19.14	22.04	12.73	150.61
SD's	10.68	4.11	8.49	7.51	6.31	57.51
Multiple R	0.631					

* Variables 1 to 5 inclusive: Sub-tests of MLAT (Predictors)
Variable 6 : MLA-LB (Criterion)

Table XVI

Intercorrelation Matrix, Means, SD's, and Multiple R for Predictor Variables (sub-tests of MLAT) for Ss Classified According to High 1/3 on Achievement Motivation (N=100)

Variables*	1	2	3	4	5	6
1	1.00	0.25	0.40	0.13	0.37	0.28
2	0.25	1.00	0.45	0.23	0.28	0.35
3	0.40	0.45	1.00	0.29	0.36	0.25
4	0.13	0.23	0.29	1.00	0.23	0.44
5	0.37	0.28	0.36	0.23	1.00	0.27
6	0.28	0.35	0.25	0.44	0.27	1.00
Means	30.64	22.89	21.48	23.25	13.15	155.19
SD's	10.04	3.90	9.33	6.62	5.73	56.57
Multiple R	0.544					

* Variables 1 to 5 inclusive: Sub-tests of MLAT (Predictors)
Variable 6 : MLA-LB (Criterion)

set of K = 3 values of R are shown in Table XVII. When sub-groups are of unequal size, Edwards[20] provides a formula that uses differentially weighted z values. Since subgroups are of equal size in this research, the formula can be simplified to become:

$$\chi^2 = (n - 3) \left[\Sigma z^2 - \frac{(\Sigma z)^2}{k} \right]$$

Substituting the values from Table XVII in the revised formula yields a χ^2 value of 2.425 which for 2 df is not statistically significant at $P < .05$. This method of comparing R's as proposed in Edwards[21] appears to be too conservative, i.e.- making significant differences difficult to find. One reason is suggested for this statement:

> By the "interocular traumatic test", i.e.- the "eye-balling method", it seems odd that the difference between R's of .632 and .482 (40% and 23% of the variances respectively) does not come out as significant.

As previously stated by Zedeck[22], this method for demonstrating differential validities, i.e.- significantly different validity coefficients among sub-groups, is more powerful than a second method, i.e.- the validity coefficients for some sub-groups are

---

20 Idem, ibid., p. 83.

21 Idem, ibid., p. 83-84.

22 Sheldon Sedeck, "Problems with the Use of Moderator Variables", in Psychological Bulletin, 1971, Vol. 76, No. 4, p. 303.

Table XVII

Calculation for the χ^2 Test of Homogeneity of k Values of R

(1) Sub-groups*	(2) n	(3) R	(4) z	(5) z^2	(6) n-3
i	100	.482	.523	.274	97
J	100	.631	.741	.549	97
K	100	.544	.604	.365	97
Σ	300	1.657	1.868	1.188	291

* i = Low 1/3 Group on Achievement Motivation
J = Middle 1/3 Group on Achievement Motivation
K = High 1/3 Group on Achievement Motivation

FORMULA (REVISED)

$$\chi^2 = (n-3)\left[\Sigma z^2 - \frac{(\Sigma z)^2}{k}\right]$$

$$= (97)\left[1.188 - \frac{(1.868)^2}{3}\right] = 2.425 \text{ (N.S)}$$

significantly different from zero, while validity coefficients for other sub-groups are not statistically significant. Accepting the fact that the above method may be conservative but also that it is one approach for demonstrating differential validity, it was concluded from the data that the three multiple R's for sub-groups I, J, and K in Table XVII are homogeneous and can be considered as estimates of the same population value. Null Hypothesis 4, therefore, is accepted.

For Null Hypothesis 5, that there is no significant difference in prediction for Ss classified according to various combinations of Integrative Orientation and Achievement Motivation levels, multiple R's were computed independently for each of the six groups. The intercorrelation matrices, means, SD's and multiple R's are shown in Tables XVIII, XIX, XX, XXI, XXII, and XXIII respectively. In comparing the multiple R's all values were transformed to Z's and the standard error of the difference was determined by the χ^2 test for homogeneity of k values of R[23], again adopting the revised formula used for testing Null Hypothesis 4.

The calculations involved in the test for a set of $k \simeq 6$ values of R are shown in Table XXIV. Substituting the values from the table in the simplified formula yields a χ^2 value of 2.773 which for 5 df is not statistically significant. Since

---

23 Idem, ibid., p. 83-84.

Table XVIII

Intercorrelation Matrix, Means, SD's, and Multiple R for Predictors (sub-tests of MLAT) and Criterion Measure (MLA-LB) for Ss Classified According to High Integrative Orientation (Split at Median) and Low 1/3 on Achievement Motivation (N = 50)

Variable*	1	2	3	4	5	6
1	1.00	0.42	0.19	0.30	0.67	0.27
2	0.42	1.00	0.45	0.39	0.23	0.49
3	0.19	0.45	1.00	0.32	0.28	0.15
4	0.30	0.39	0.32	1.00	0.40	0.45
5	0.67	0.23	0.28	0.40	1.00	0.27
6	0.27	0.49	0.15	0.45	0.27	1.00
Means	30.66	22.32	17.38	19.94	13.00	127.26
SD's	9.73	4.16	9.18	7.93	4.76	56.65
Multiple R	0.590					

* Variables 1 to 5 inclusive: Sub-tests of MLAT (Predictors)
Variable 6 : MLA-LB (Criterion Measure)

Table XIX

Intercorrelation Matrix, Means, SD's, and Multiple R for Predictors (sub-tests of MLAT) and Criterion Measure (MLA-LB) for Ss Classified According to High Integrative Orientation (Split at Median and Middle 1/3 on Achievement Motivation (N = 50)

Variable*	1	2	3	4	5	6
1	1.00	0.46	0.43	0.33	0.42	0.42
2	0.46	1.00	0.56	0.53	0.38	0.46
3	0.43	0.56	1.00	0.34	0.47	0.48
4	0.33	0.53	0.34	1.00	0.37	0.47
5	0.42	0.38	0.47	0.37	1.00	0.45
6	0.42	0.46	0.48	0.47	0.45	1.00
Means	31.24	21.78	18.62	20.90	12.40	149.48
SD's	10.17	3.91	9.16	7.95	6.51	57.32
Multiple R	0.622					

* Variables 1 to 5 inclusive: Sub-tests of MLAT (Predictors)
Variable 6 : MLA-LB (Criterion Measure)

Table XX

Intercorrelation Matrix, Means, SD's, and Multiple R for Predictors (sub-tests of MLAT) and Criterion Measure (MLA-LB) for Ss Classified According to High Integrative Orientation (Split at Median) and High 1/3 on Achievement Motivation (N = 50)

Variable*	1	2	3	4	5	6
1	1.00	0.32	0.32	0.24	0.34	0.27
2	0.32	1.00	0.44	0.25	0.44	0.30
3	0.32	0.44	1.00	0.26	0.47	0.26
4	0.24	0.25	0.26	1.00	0.41	0.57
5	0.34	0.44	0.47	0.41	1.00	0.37
6	0.27	0.30	0.26	0.57	0.37	1.00
Means	30.06	22.90	21.02	21.96	13.44	155.74
SD's	10.33	3.79	8.79	6.89	5.76	52.36
Multiple R	0.604					

* Variables 1 to 5 inclusive: Sub-tests of MLAT (Predictors)
Variable 6 : MLA-LB (Criterion Measure)

Table XXI

Intercorrelation Matrix, Means, SD's, and Multiple R for Predictors (sub-tests of MLAT) and Criterion Measure (MLA-LB) for Ss Classified According to Low Integrative Orientation (Split at Median) and Low 1/3 on Achievement Motivation (N = 50)

Variable*	1	2	3	4	5	6
1	1.00	0.22	0.22	0.40	0.24	0.11
2	0.22	1.00	0.40	0.34	0.13	0.31
3	0.22	0.40	1.00	0.32	0.32	0.31
4	0.40	0.34	0.32	1.00	0.41	0.33
5	0.24	0.13	0.32	0.41	1.00	0.09
6	0.11	0.31	0.31	0.33	0.09	1.00
Means	31.12	22.68	18.06	20.86	12.06	132.76
SD's	10.02	4.25	8.57	7.90	5.42	54.31
Multiple R	0.432					

* Variables 1 to 5 inclusive: Sub-tests of MLAT (Predictors)
Variable 6 : MLA-LB (Criterion Measure)

Table XXII

Intercorrelation Matrix, Means, SD's, and Multiple R for Predictors (sub-tests of MLAT) and Criterion Measure (MLA-LB) for Ss Classified According to Low Integrative Orientation (Split at Median) and Low 1/3 on Achievement Motivation (N = 50)

Variable*	1	2	3	4	5	6
1	1.00	0.40	0.13	0.53	0.40	0.44
2	0.40	1.00	0.33	0.20	0.27	0.26
3	0.13	0.33	1.00	0.38	0.35	0.41
4	0.53	0.20	0.38	1.00	0.42	0.58
5	0.40	0.27	0.35	0.42	1.00	0.27
6	0.44	0.26	0.41	0.58	0.27	1.00
Means	29.16	22.26	19.22	22.74	12.88	148.86
SD's	11.16	4.32	7.68	6.73	5.96	56.74
Multiple R	0.643					

* Variables 1 to 5 inclusive: Sub-tests of MLAT (Predictors)
Variable 6 : MLA-LB (Criterion Measure)

Table XXIII

Intercorrelation Matrix, Means, SD's, and Multiple R for Predictors (sub-tests of MLAT) and Criterion Measure (MLA-LB) for Ss Classified According to Low Integrative Orientation (Split at Median) and High 1/3 on Achievement Motivation (N = 50)

Variable*	1	2	3	4	5	6
1	1.00	0.17	0.48	0.02	0.42	0.29
2	0.17	1.00	0.47	0.23	0.15	0.39
3	0.48	0.47	1.00	0.31	0.32	0.26
4	0.02	0.23	0.31	1.00	0.13	0.37
5	0.42	0.15	0.32	0.13	1.00	0.24
6	0.29	0.39	0.26	0.37	0.24	1.00
Means	31.48	22.90	22.38	24.98	13.04	157.52
SD's	9.80	4.05	9.90	6.24	5.96	62.16
Multiple R	0.563					

* Variables 1 to 5 inclusive: Sub-tests of MLAT (Predictors)
Variable 6 : MLA-LB (Criterion Measure)

Table XXIV

Calculations for the χ^2 Test of Homogeneity of k Values of R

(1)* Subgroups	(2) N	(3) R	(4) z	(5) z^2	(6) n-3
A	50	.590	.678	.460	47
B	50	.622	.725	.526	47
C	50	.604	.701	.491	47
D	50	.432	.460	.212	47
E	50	.643	.767	.588	47
F	50	.563	.640	.410	47
Σ	300	3.454	3.971	2.687	282

* A= High Group on Integrative Orientation, Low 1/3 on Achievement Motivation
B= High Group on Integrative Orientation, Middle 1/3 on Achievement Motivation
C= High Group on Integrative Orientation, High 1/3 on Achievement Motivation
D= Low Group on Integrative Orientation, Low 1/3 on Achievement Motivation
E= Low Group on Integrative Orientation, Middle 1/3 on Achievement Motivation
F= Low Group on Integrative Orientation, High 1/3 on Achievement Motivation

FORMULA (REVISED)

$$\chi^2 = (n-3)\left[\Sigma z^2 - \frac{(\Sigma z)^2}{k}\right]$$

$$(47)\left[2.687 - \frac{(3.971)^2}{6}\right] = 2.773 \quad \text{(N.S.)}$$

the χ^2 test did not result in a rejection of the null hypothesis it was considered desirable to obtain a more exact estimate of the population correlation using Fisher's correction term for the bias[24]. The value of χ^2 was re-computed in Table XXV. From Table XXIV the average of z is .66. The R corresponding to this value is .58. This value of .58 was regarded as an estimate of the common population correlation based upon the data under consideration. The corrected z for the first sample in Table XXV was:

$$\text{Corrected } z = .678 - \frac{.58}{2\ (50-1)} = .672$$

The corrected z values for other five samples are .719, .695, .454, .761, and .634 respectively. These corrected values are reported in Column 4 of Table XXV.

The correction term for the z values (see Table XXV) was found to be of relatively little importance in this case with an N of 50 in each sample and a total of 6 samples. Again, substituting the values in Table XXV in the simplified formula, yields a χ^2 value of 2.726 which for 5 df is not statistically significant for $P < .05$. Again as for Null Hypothesis 4 it would appear that the χ^2 test may be extremely conservative. By visual inspection the multiple R's range from

---

24 Idem, ibid., p. 84.

Table XXV

Calculations for the χ^2 Test of Homogeneity of k Values of R Using Fisher's Correction Term for the Bias

(1)* Subgroups	(2) N	(3) R	(4) z (corrected)	(5) z^2	(6) n-3
A	50	.590	.672	.452	47
B	50	.622	.719	.517	47
C	50	.604	.695	.483	47
D	50	.432	.454	.206	47
E	50	.643	.761	.579	47
F	50	.563	.634	.402	47
Σ	300	3.454	3.935	2.639	282

* A= High Group on Integrative Orientation, Low 1/3 on Achievement Motivation
B= High Group on Integrative Orientation, Middle 1/3 on Achievement Motivation
C= High Group on Integrative Orientation, High 1/3 on Achievement Motivation
D= Low Group on Integrative Orientation, Low 1/3 on Achievement Motivation
E= Low Group on Integrative Orientation, Middle 1/3 on Achievement Motivation
F= Low Group on Integrative Orientation, High 1/3 on Achievement Motivation

FORMULA (REVISED)

$$\chi^2 = (n-3)\left[\Sigma z^2 - \frac{(\Sigma z)^2}{k}\right]$$

$$= (47)\left[2.639 - \frac{(3.935)^2}{6}\right] = 2.726$$

(N.S.)

.432 to .643 (19% and 41% of the variances respectively) but do not come out as significantly different. Nevertheless, from the statistical test it was concluded that the six multiple R's for subgroups A to F conclusive are homogeneous and can be considered as estimates of the same population value. Null Hypothesis 5, therefore, is accepted.

Prior to presenting data relevant to the testing of null hypotheses 6 to 9 inclusive, a brief summary of the statistical procedures used up to this point in this research to demonstrate differential validities is perhaps in order. Three methods cited by Zedeck[25] have been used by different investigators for demonstrating differential validities. The first method, comparing correlations among sub-groups, was the statistical method used to test null hypotheses 3 to 5 inclusive. This method is suggested by Zedeck[26] to be more powerful than a second method: using patterns of validity coefficients that reveal some coefficients that are significantly different from zero. The only meaningful use made of this latter statistic in this research was to test for statistical significance of validity coefficients on cross-validation. This method was not therefore used to demonstrate differential validities. A third method cited was testing whether the

---

25 Zedeck, Op. Cit., p. 303.

26 Idem, ibid., p. 303.

validity coefficient of a subgroup is significantly different from the validity coefficient of the total group. According to Zedeck[27] this would be the most desirable method to determine whether moderators are adding to prediction. Yet Zedeck[28] states that he does not know of a statistic that tests for this although different investigators formed conclusions based on such analyses.

As noted in the chapter on Experimental Design a formula suggested by Dayhaw[29] is used to compare the validity coefficient of a subgroup to the validity coefficient of a total group. This formula,

$$\sigma R's = \frac{1 - R^2t}{\sqrt{Ns - 1}} \sqrt{\frac{Nt - Ns}{Nt - 1}}$$

where

Rt = Multiple R for total group;
Nt = Number in total group;
Ns = Number in the subgroup

is suggested as the standard error of a difference between a multiple correlation coefficient of a subgroup and the corresponding multiple correlation coefficient in the total group which contains the subgroup. The above formula will subsequently be used to test null hypotheses 6 to 9 inclusive.

---

27 Idem, ibid., p. 303.

28 Idem, ibid. p. 303.

29 In a personal communication.

For Null Hypothesis 6 which required a comparison of validity coefficients computed in subgroups and the validity coefficients computed for the total group, a summary of relevant data is reported in Table XXVI. Applying the formula suggested by Dayhaw[30] to these data yielded standard errors of difference values as follows: Subgroups A, B, C, D, E and F, each compared to the total group (T) = .0983; Subgroups G and H, each compared to T = .0412; Subgroups I, J, and K, each compared to T = .0583. Table XXVII presents the R's, z's, Diff/σR's and levels of statistical significance for the sub-group and total group comparisons of multiple R's reported in Table XXVI.

The data in Tables XXVI and XXVII show multiple R's ranging from .432 to .643 for the different subgroups. In other words the percentage of the variance in the criterion (Y) associated with the variance in the predictors (X's) ranges from 19% to 41% respectively. For the total group (T) the corresponding value is 29%. Viewing the data by the "interocular traumatic test", i.e., the "eye-balling method", eight of the eleven subgroups have numerically higher multiple R's than the total group. From a statistical viewpoint only subgroup J (High 1/3 on Achievement Motivation) has a significantly higher multiple R than the total group

---

30 Idem, ibid.

Table XXVI

Multiple R's Between Sub-Tests of MLAT (Predictors) and the MLA-LB (Criterion Measure) for Ss Classified According to Three Achievement Motivation Levels and Two Integrative Orientation Levels

	Achievement Motivation			
	Low 1/3 Group	Middle 1/3 Group	High 1/3 Group	Total Group
	(a)	(b)	(c)	(g)
High Group	R: 0.590 (N = 50)	R: 0.622 (N = 50)	R: 0.604 (N = 50)	R: 0.570 (N=150)
	(d)	(e)	(f)	(h)
Low Group	R: 0.432 (N = 50)	R: 0.643 (N = 50)	R: 0.563 (N = 50)	R: 0.516 (N=150)
	(i)	(j)	(k)	(t)
	R: 0.482 (N = 100)	R: 0.631 (N = 100)	R: 0.544 (N = 100)	R: 0.539 (N = 300)

Table XXVII

Multiple R's, z's, Diff/σRs, and Level of Statistical Significance for Subgroup and Total Group Comparisons of Multiple R's Reported in Table XXVI

Groups *	R	z	$\frac{\text{Diff.}}{\sigma R's}$	P
Total Group (T)	.539	.604		
A	.590	.678	.753	NS
B	.622	.725	1.231	NS
C	.604	.693	.905	NS
D	.432	.460	-1.465	NS
E	.643	.758	1.567	NS
F	.563	.633	.295	NS
G	.570	.648	1.067	NS
H	.516	.577	- .655	NS
I	.482	.523	-1.389	NS
J	.631	.741	2.350	S ($P < .05$)
K	.544	.611	.120	NS

* Groups Defined as per Table XXVI

($P < .05$) although subgroups D and E approach statistical significance (lower and higher respectively). Again the fact that more of the comparisons did not reach the level of statistical significance could be explained by the decreasing sample sizes due to subgrouping. This is cited by Zedeck[31] as one of the problems that plagues the investigator in subgroup analyses. For Null Hypothesis 6, therefore, that there is no significant difference between the multiple R of each subgroup and the multiple R of the total group, the null hypothesis is accepted for subgroups A to I inclusive and subgroup K, but is rejected for subgroup J ($P < .05$).

To test Null Hypothesis 7, multiple R's computed for sub-groups G and H of Table XXVI were compared to multiple R for the total group (T). The formula suggested by Dayhaw[32] was again used (as for Null Hypothesis 6) except that in computing the multiple R for the total group the Integrative Orientation measure was entered into the regression equation as one of the predictor variables. The multiple R's for sub-groups G and H (Table XXVII) are .570 and .516 respectively (z's = .648, and .577), and the corresponding multiple R for the total group with Integrative Orientation included as a predictor variable in addition to

---

31 Zedeck, Op. Cit., p. 303.

32 Dayhaw, Op. Cit.

the MLAT subtests (See Table XI), is .540 ($z = .604$). Computing σ_R, a value of .0411 is obtained. This value is observed to be identical to that obtained in comparing subgroups G and H to the total group in testing Null Hypothesis 6. The finding is not unexpected in that it was initially found that the inclusion of the Integrative Orientation predictor did not significantly improve prediction when added to the MLAT subtests in the multiple regression equation (Null Hypothesis 2). To test whether Integrative Orientation functions better as a moderator than as a predictor, then $z_1 - z_2/\sigma_R$

For subgroup G: $$\frac{.648 - .604}{.0411} = 1.063 \text{ (NS)}$$

For subgroup H: $$\frac{.577 - .604}{.0411} = -.657 \text{ (NS)}$$

It can be noted that the matter of whether σ_R should be converted into σ_z in this case is purely academic since at this level of correlation of .04, $R = z$. Null Hypothesis 7 is therefore accepted. No significant improvement has been found when Integrative Orientation is used as a moderator over it being used as a predictor.

In testing Null Hypothesis 8, multiple R's computed for subgroups I, J, and K (See Table XXVII) were compared to the multiple R for the total group (T). For the total group

the Achievement Motivation measure was entered in the regression equation as one of the predictor variables in addition to the MLAT subtests. The multiple R's for subgroups I, J, and K are .482, .631, and .544 respectively (z's = .523, .741 and .611) and the corresponding multiple R for the total group (T) with the Achievement Motivation variable included as a predictor variable (See Table XI) is .546 (z = .611). Computing σR[33], a value of .0577 is obtained.

As previously noted (Null Hypothesis 2), the inclusion of the Achievement Motivation measure did not significantly improve prediction when added to MLAT sub-tests in a multiple regression equation. To test whether Achievement Motivation functions better as a moderator than as a predictor, then $z_1 - z_2 / \sigma R$

For subgroup I: $\frac{.523 - .611}{.0577} = -1.525$ (NS)

For subgroup J: $\frac{.741 - .611}{.0577} = 2.253$ (S)

For subgroup K: $\frac{.611 - .611}{.0577} = .000$ (NS).

---

33 Idem, ibid.

Null Hypothesis 8 is therefore accepted for subgroups I and K, but is rejected for subgroup J ($P < .05$). In other words, for the Middle 1/3 Achievement Motivation group it was found that the motivation measure functioned better as a moderator than as a predictor in a multiple regression model. For the Low 1/3 and High 1/3 Achievement Motivation groups similar comparisons did not yield significant results.

To test Null Hypothesis 9, multiple R's computed for subgroups A to F inclusive (See Table XXVI) were compared to the multiple R for the total group (T). For the total group, both the Integrative Orientation and Achievement Motivation variables were included as predictors in the multiple regression equation in addition to the MLAT sub-tests. From Table XXVI the multiple R's for subgroups A to F inclusive are .590, .622, .604, .432, .643 and .563 respectively (z's = .678, .725, .693, .460, .758, and .633). The multiple R for the total group (T) with both the orientation and motivation variables included as predictor variables in addition to the aptitude sub-tests (See Table XV) is .547 (z = .618). Computing σ_R[34] for these data yielded a value of .0929. To test whether both measures of Integrative Orientation and Achievement Motivation function better as moderators than as predictors in a multiple regression equation all multiple R's were transformed to z's and (D/σ_R) obtained in the usual manner.

---

34 Idem, ibid.

The results of these analyses are reported in Table XXVIII. From the data, no significant improvement has been found when both Integrative Orientation and Achievement Motivation variables are used as moderators over both variables being used as predictors. Null Hypothesis 9 is therefore accepted. Trends to statistical significance are noted in subgroups D and E (one lower and one higher respectively).

To test Null Hypothesis 10 that there is no significant difference in performance in intensive language training for Ss classified according to: High/Low Groups on MLAT; High 1/3, Middle 1/3, and Low 1/3 Groups on Achievement Motivation; and High/Low Groups on Integrative Orientation, a 2 X 3 X 2 factorial analysis of variance design[35] with unequal cell frequencies was utilized. Table XXIX presents the results of this analysis. The performance scores (MLA-LB), N's, Means, SD's and variances for each of the 12 groups so constituted are recorded in Appendix 4. Since the N's varied for the different sub-groups, Hartley's[36] F Max statistic was applied to the data in Appendix 4 to test for homogeneity of variance as follows:

$$F\ Max = \frac{\text{Largest of k variances}}{\text{Smallest of k variances}} = \frac{4257.618}{1893.970} = 2.248.$$

---

35 BMDX64 -- General Least Squares Method -- Revised January 20, 1969, Health Sciences Computing Facility, UCLA.

36 Roger E. Kirk, Experimental Design: Procedures for the Behavioral Sciences, Brooks/Cole Publishing Co., Belmont, California, USA, 1968, p. 62.

Table XXVIII

Multiple R's, z's, D/$\sigma_{\bar{R}}$ and Levels of Statistical Significance for Subgroup and Total Group Comparisons of Multiple R's Required for Testing of Null Hypothesis 9 ($\sigma_{\bar{R}} = .0929$).

Groups	R	z	$\frac{\text{Diff.}}{\sigma_{\bar{R}}}$	P
Total Group (T)	.547	.618		
A	.590	.678	.646	NS
B	.622	.725	1.152	NS
C	.604	.693	.807	NS
D	.432	.460	- 1.701	NS
E	.643	.758	1.507	NS
F	.563	.633	.161	NS

Table XXIX

Analysis of Variance Summary Table: MLA-LB Performance Scores in Intensive Language Training Under A: Two Levels of Language Aptitude (High/Low); B: Three Levels of Achievement Motivation (High/Middle and Low); and C: Two Levels of Integrative Orientation

Source	MLA-LB Criterion of Performance			
	SS	df	MS	F
A	153705.56	1	153705.56	57.6235**(a)
B	27922.49	2	13961.24	5.2340**(b)
C	6.12	1	6.12	0.0023
AB	8373.96	2	4186.98	1.5697 (c)
AC	0.31	1	0.31	0.0001
BC	2279.96	2	1139.98	0.4274
ABC	3827.86	2	1913.93	0.7175
Error	768214.88	288	2667.41	
Total	964331.14	299		

** (a) Significance at $P < .01$ requires an F value above 6.76.

** (b) Significance at $P < .01$ requires an F value above 4.71.

(c) Significance at $P < .05$ requires an F value above 3.04.

With df = k and n-1, where k = the number of variances (here 12) and n is the number of observations within each level (here approximately 25), an F max value of about 3.39 would be required for $P < .05$. It is concluded therefore that the variances can be considered as homogeneous.

The data in Table XXIX show that the main effects for ability and motivation have both been found to be significant ($P < .01$), while the main effect for orientation has not reached the level of statistical significance. In view of the earlier findings in this research with the multiple regression model, the highly significant effect of ability and of motivation on performance was to be expected. However, caution must be exercised in interpreting the main effects when the bases of classification are subject status variables. As pointed out by McNemar[37], such variables are unlikely to be totally independent. In this research the analysis of variance is used primarily to test for interactive or multiplicative effects of ability, motivation, and orientation on performance in intensive language learning. In order to compare results of this research with Lawler's[38] findings, it was considered necessary to adopt the same procedure rather than utilizing a

---

37 McNemar, Op. Cit., p. 445-446.

38 Edward E. Lawler III, "Ability as a Moderator of the Relationship Between Job Attitudes and Job Performance", in Personnel Psychology, Vol. 19, 1966, p. 153-164.

moderated multiple regression model. Consistent with Vroom's[39] formulation, the interaction of ability and motivation on performance is specified as multiplicative. Another consideration in using ANOVA was Zedeck's[40] general comment that moderated multiple regression has not been successful in improving predictions. Again referring to Table XXIX, the interaction effects of ability, motivation, and orientation on performance were computed and not found to be statistically significant. Null Hypothesis 10, therefore, is accepted. No support has been found in this research for the interactive or multiplicative model. In the event that the finding of insignificant interactive effects might be due to the small numbers in the different subgroups the performance means are plotted graphically in Figure 1 so that at least trends to statistical significance might be observed.

Figure 1 shows rather similar trends for all subgroups represented by A_1C_1, A_1C_2 and A_2C_1. An inconsistency that would appear difficult to interpret is one represented by A_2C_2. The subgroup classified as Low Aptitude/Middle Motivation/Low Orientation performs more poorly than the subgroup designated as Low Aptitude/Low Motivation/Low Orientation

---

39 Vroom, Op. Cit., p. 331.

40 Zedeck, Op. Cit., p. 302.

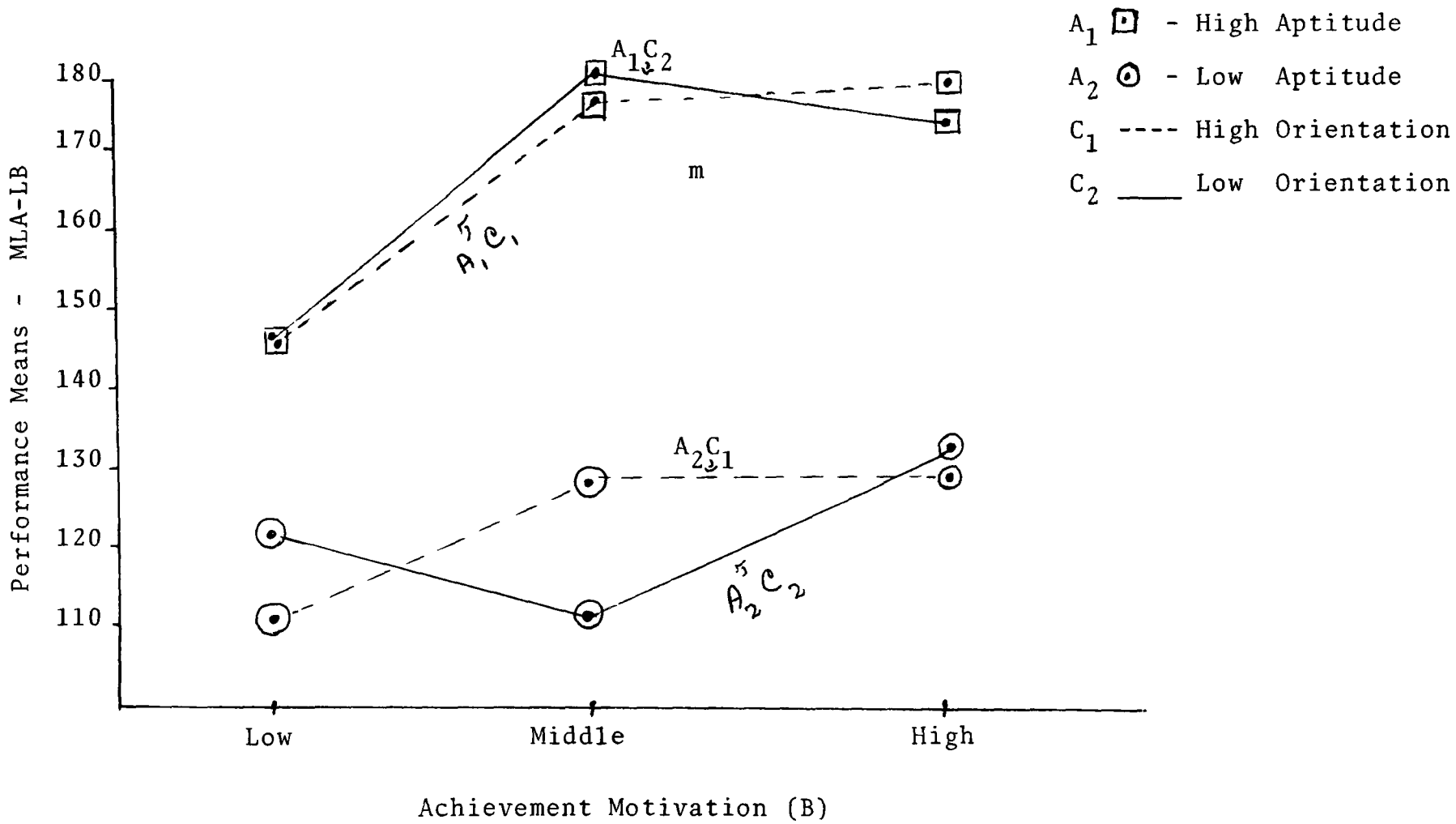


Figure 1. MLA-LB Performance Criterion Means for Ss Classified Under A: Two Levels of Language Aptitude; B: Three Levels of Achievement Motivation; and C: Two Levels of Integrative Orientation.

Again the greatest difference between the performance means would seem to be between the subgroups classified as: Low Aptitude/Middle Motivation/High Orientation; and Low Aptitude/ Middle Motivation/Low Orientation. Although it could be stated that there is a trend to interaction, it must be remembered that the last two subgroups discussed consist of 28 and 21 Ss respectively and the observed differences could be due to random fluctuations.

As an additional precaution, although the robustness of the F test under the usual conditions[41] is acknowledged, Ss were drawn at random from the different subgroups so that all groups were of equal size, and the analysis of variance recomputed. The results of this analysis is presented in Table XXX. The main effects for both aptitude and motivation on performance are again statistically significant while the main effect for orientation is not significant. The interactions are not statistically significant although a trend to statistical significance is noted for the (Aptitude-Motivation) A:B interaction.

---

41 G.E.P. Box, "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems", in the Annals of Mathematical Statistics, 25, p. 290-302.

Table XXX

Analysis of Variance Summary Table: MLA-LB Performance Scores in Intensive Language Learning Under A: Two Levels of Language Aptitude (High/Low); B: Three Levels of Achievement Motivation (High/Middle/Low); and C: Two Levels of Integrative Orientation (High/Low) - Ss Selected Out at Random to Equate All Groups for Size at n = 21.

Source	Ss	df	Ms	F
A	113496.45	1	113496.45	39.1072**(a)
B	25073.92	2	12536.96	4.3198*(b)
C	464.15	1	464.15	.1599
AB	13173.90	2	6586.95	2.2696 (c)
AC	57.14	1	57.14	.0197
BC	2777.16	2	1388.58	.4784
ABC	2547.94	2	1273.97	.4390
Error	696525.34	240	2902.19	
Total	854116.00	251		

**(a) Significance at $P < .01$ requires an F value above 6.76

*(b) Significance at $P < .05$ requires an F value above 3.04

(c) Significance at $P < .05$ requires an F value above 3.04

5. Subsequent Analyses

As suggested by Zedeck[42], a statistical problem pertinent to all moderator approaches is that of cross-validation. Since in this research, improved prediction has been demonstrated for a middle motivation subgroup when a measure of Achievement Motivation is used as a moderator over it being used as a predictor (Null Hypothesis 8) it was considered mandatory to cross-validate for subgroups designated as I, J, and K (See Table 1 of Chapter 1). For each subgroup of 100 Ss, (High 1/3, Middle 1/3, Low 1/3), using a table of random digits, two random samples of 50 Ss each (A and B) were selected.

For each sample (A and B), optimum weights were obtained and a multiple R computed. The regression weights derived for one sample were then applied to the scores for the other sample. In other words, a double cross-validation design was utilized. The results of these analyses are recorded in Table XXXI.

The multiple R's for the different groups range from .348 to .704. The cross-validation R's range from .260 to .632. In evaluating the R's in Table XXXI for purposes of comparison the relevant R's are the cross-validation R's. In

---

42 Idem, ibid., p. 302.

Table XXXI

Multiple R's and Cross-Validation R's for Subtests of MLAT (Predictors) and MLA-LB (Criterion) for Ss Classified as High/Middle/Low 1/3 Groups on Achievement Motivation (Moderator)

Subgroup #		N	Multiple R	Cross-Validation R		"Drop" in R
Low Motivation	A	50	.348	A to B	.260	.088
	B	50	.704**	B to A	.536**	.168
Middle Motivation	A	50	.640**	A to B	.632**	.008
	B	50	.627**	B to A	.618**	.009
High Motivation	A	50	.587**	A to B	.340	.247
	B	50	.623**	B to A	.411	.212

Low/Middle/High 1/3 Groups on Achievement Motivation, each subgroup further divided randomly into two equal Groups (A and B)

** Significance at $P < .01$ requires an R value $\geq .479$

* Significance at $P < .05$ requires an R value $\geq .412$

discussing the approaches used by different investigators to demonstrate differential validity, Zedeck[43] evaluates three different approaches. One approach -- patterns of validity coefficients reveal some coefficients to be significantly different from zero -- does not merit much comment. From Table XXXI the cross-validation R for Low Motivation subgroup B is significantly different from zero while for Low Motivation subgroup A the cross-validation R is not statistically significant. For both Middle Motivation subgroups A and B the cross-validation R's are statistically significant ($P < .01$). The cross-validation R's for both High Motivation subgroups A and B are not statistically significant. Perhaps a more relevant comment here is the observation in terms of the "drop" in R on cross-validation. For both middle motivation subgroups the "drop" is insignificant (using the "eye-balling test"). On the other hand, the "drop" in cross-validation R's for High Motivation subgroup A and B is substantial; this is also observed to be the case for Low Motivation subgroup B.

Using a second approach suggested by Zedeck[44] to be more powerful than the first approach, i.e., the R for one subgroup is significantly different from the R for another subgroup, all cross-validation R's were compared using the χ^2

---

43 Idem, ibid., p. 302.

44 Idem, ibid., p. 302.

Test[45] for homogeneity of k values of R. The data required for calculation using a revised formula (since the subgroups are of equal size) are presented in Table XXXII. The χ^2 value obtained was 9.212 which for 5 df is not significant. Since this value was not found to be statistically significant, a correction for bias was made and the results are presented in Table XXXIII. With six subgroups of 50 Ss each the correction for bias was not found to have a significant effect.

Again, referring to the data in Tables XXXII and XXXIII, it seems odd that the difference between cross-validation R's of .260 and .632 is not statistically significant (.068 and .399 of the variances respectively). Significance at $P < .05$ with K-1 df, or 5df would have required a χ^2 value of 11.07. It is repeated that the χ^2 test would appear to be extremely conservative for the purpose of demonstrating differential validity.

A third method is suggested by Zedeck[46] to be the most desirable to determine whether moderators are adding to prediction, i.e., the validity coefficient of a subgroup is different from the validity coefficient of the total group. As a tentative measure, the formula suggested by Dayhaw[47]:

---

45 Edwards, Op. Cit., p. 82-83.

46 Zedeck, Op. Cit., p. 302.

47 Dayhaw, Op. Cit.

Table XXXII

Calculations for $\chi^{2(1)}$ Test of Homogeneity of k Values of Cross-Validated R's

(1) Subgroups*	(2) N	(3) R	(4) z	(5) z^2	(6) n-3
1	50	.260	.266	.071	47
2	50	.536	.604	.364	47
3	50	.632	.741	.549	47
4	50	.618	.725	.526	47
5	50	.340	.354	.125	47
6	50	.411	.436	.190	47
Σ	300	2.797	3.126	1.825	282

* Subgroups as defined in Table XXXI

(1) Amended Formula:

$$\chi^2 = (n-3)\left[\Sigma z^2 - \frac{(\Sigma z)^2}{k}\right]$$

$$= (47)\ [1.825-1.629]$$

$$= 9.212 \quad (N.S.)$$

Table XXXIII

Calculation of the χ^2 (1) Test of Homogeneity of k Values of R Using Fisher's (2) Correction Term for the Bias

(1) Subgroups*	(2) N	(3) R	(4) z (corrected)	(5) z^2	(6) n-3
1	50	.260	.261	.068	47
2	50	.536	.599	.359	47
3	50	.632	.736	.542	47
4	50	.618	.720	.518	47
5	50	.340	.349	.122	47
6	50	.411	.431	.186	47
Σ	300	2.797	3.096	1.795	282

* Subgroups as defined in Table XXXI

(1) Amended Formula: $\chi^2 = (n-3)\left[\Sigma z^2 - \frac{(\Sigma z)^2}{k}\right]$

$= (47)\left[1.795 - \frac{(3.096)^2}{6}\right]$

$= 9.259$ (N.S)

(2) Correction Term $= \frac{P}{2(n-1)}$ (Subtracted from each z value)

$$\sigma_{R's} = \frac{1 - R^2_t}{\sqrt{N_s - 1}} \sqrt{\frac{N_t - N_s}{N_t - 1}}$$

will again be used to determine the standard error of the difference between a cross-validation R of a subgroup (Ns) and the corresponding cross-validation R in the total group (N_t).

For the total group the best estimate of the cross-validation R with both motivation and orientation included in addition to the MLAT subtests would seem to be the mean cross-validation R's for samples A and B Full Model (See Table X). These values are .409 and .476 respectively - (z's of .436 and .517). Summing and dividing by 2 yields a $\bar{z}$ value of .477 ($\bar{R} = .44$).

The best estimates of the cross-validation R's with motivation used as a moderator would seem to be the mean cross-validation R's for each of the Low, Middle and High Motivation subgroups (See Table XXXI). For: (a) - the Low Motivation groups, the cross-validation R's are .260 and .536 (z's = .266 and .604 respectively); for which $\bar{z} = .435$ ($\bar{R} = .41$). For (b) - the Middle Motivation groups, the cross-validation R's are .632 and .618 (z's = .741 and .725 respectively), for which $\bar{z} = .733$. ($\bar{R} = .63$). The corresponding cross-validation R's for (c) - the High Motivation groups, are .340 and .411 (z's = .354 and .436 respectively); for which $\bar{z} = .395$ ($\bar{R} = .38$). Applying the appropriate values in the formula, for (a), (b),

and (c): $\sigma_R = \frac{1 - (.44)^2}{\sqrt{100 - 1}} \sqrt{\frac{300 - 100}{300 - 1}}$

$= .0663$

then, for (a) - Low Motivation Subgroup to Total Group

$= \frac{.435 - .477}{.0663} = - 633$ (NS);

for (b) - Middle Motivation Subgroup to Total Group

$= \frac{.733 - .477}{.0663} = - 3.861$ (S);

and for (c) - High Motivation Subgroup to Total Group

$= \frac{.395 - .477}{.0663} = - 1.237$ (NS).

From the above data it was found that there was no significant difference in prediction when cross-validated R's for Low and High Motivation subgroups (Motivation as moderator) are compared to cross-validation R's for the total group (Motivation used as a predictor). When cross-validation R's for a Middle Motivation sub-group (Motivation as a moderator) are compared to cross-validation R's for the total group (Motivation included as a predictor) a significant improvement in prediction has been found in favor of motivation used as a moderator ($P < .01$).

Since a significant interaction was not found (Null Hypothesis 10) for either the two factor interactions: Aptitude and Motivation (AB), Aptitude and Orientation (AC), and Motivation and Orientation (BC); or the three factor interaction (ABC) in Table XXIV, and the three factor interaction is insignificant

at the .25 level[48], some pooling of groups was effected. The Orientation variable was dropped and the 12 groups were re-categorized into six groups. The analysis of variance data for these groups are presented in Table XXXIV. Again the main effects for both aptitude and motivation are highly significant ($P < .01$), but the AB interaction has not been found to be statistically significant. A trend to statistical significance is noted. The performance means (MLA-LB) have been plotted graphically in Figure 2. Since in contrast to data previously plotted (Figure 1) these data are based on performance means with larger N's, greater stability is assured. It is indicated ("eye-balling test") that for the High Aptitude subgroups, the subgroup designated as High Aptitude/Middle Motivation performs better than either the High Aptitude/Low Motivation subgroup or the High Aptitude/High Motivation subgroup. This same trend is not indicated for the Low Aptitude subgroups. If these different trends had been significant, a significant interaction would have been expected.

---

48 Roger E. Kirk, Experimental Design for the Behavioral Sciences, Brooks/Cole Publishing Company, Belmont, U.S.A., 1968, p. 214-215.

Table XXXIV

Analysis of Variance Summary Table: MLA-LB Performance Scores in Intensive Language Learning Under A: Two Levels of Language Aptitude (High/Low), and B: Three Levels of Achievement Motivation (High/Middle/Low)

Source	Ss	df	Ms	F
A	153705.56	1	153705.56	58.3595** (a)
B	27922.49	2	13961.24	5.3009** (b)
AB	8373.96	2	4186.98	1.5897 (c)
Error	774329.13	294	2633.77	
Total	964331.14	299		

** (a) Significance at $P < .01$ requires an F value above 6.76

** (b) Significance at $P < .01$ requires an F value above 4.71

(c) Significance at $P < .05$ would require an F value above 3.04

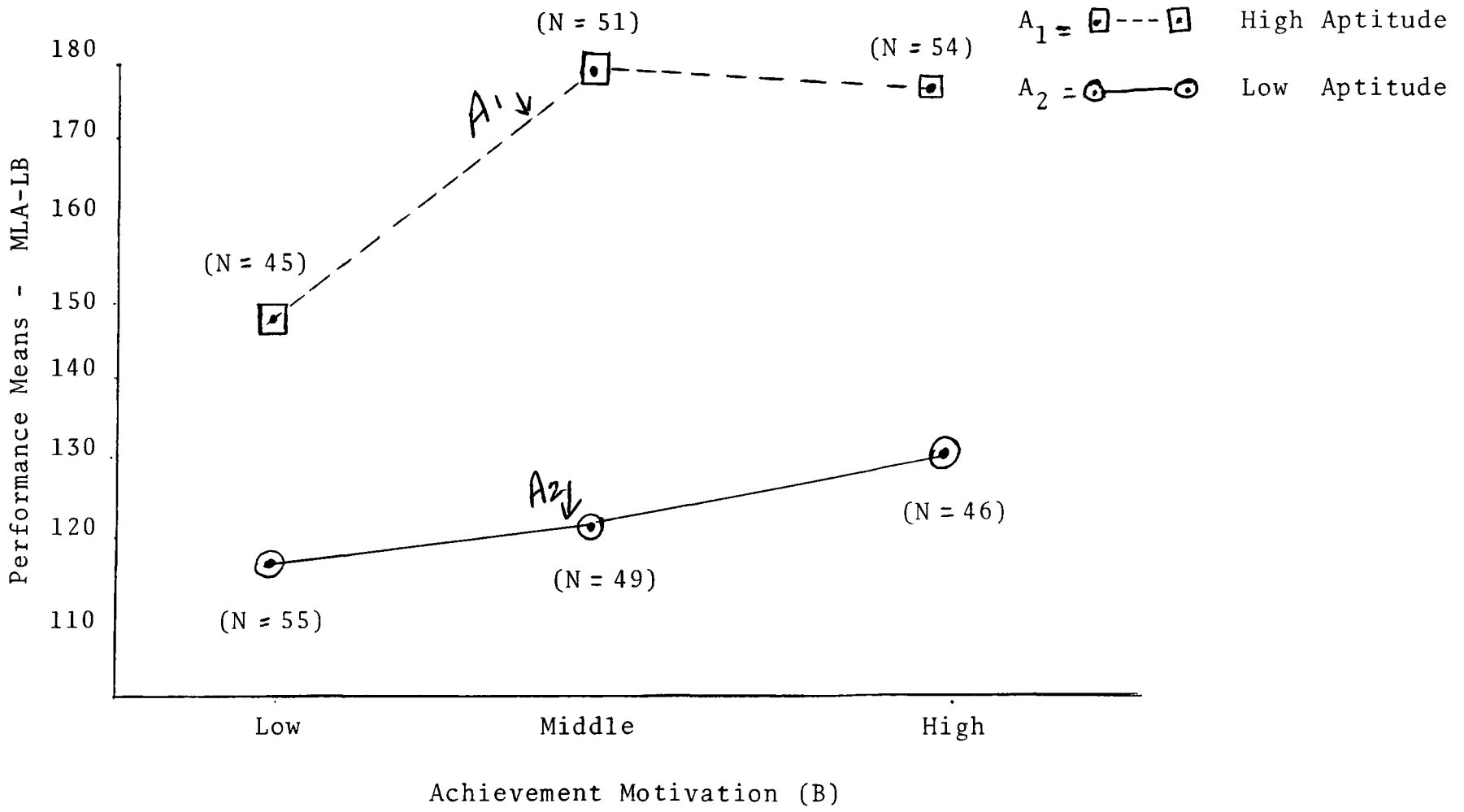


Figure 2. MLA-LB Performance Means for Ss Classified Under A: Two Levels of Language Aptitude; and B: Three Levels of Achievement Motivation.

SUMMARY AND CONCLUSIONS

This summary of research findings in the context of the Canadian Public Service intensive French language training program supports the following general conclusions:

1. In a first validation sample, no significant difference was found between the cross-validation R for a Full Multiple Regression Model using all predictors (aptitude sub-tests, motivation and orientation) and the corresponding cross-validation R for a Stepwise Model using three predictors only (R's of .409 and .410 respectively). For a second validation sample, the cross-validation R for the Full Multiple Regression Model was found to be significantly higher ($P < .01$) than the cross-validation R for the Stepwise Model (.476 and .336 respectively).

2. For the Full Multiple Regression Model paradigm, the inclusion of the motivation and orientation variables as predictors did not significantly increase validity over the aptitude sub-tests used alone (multiple R's of .547 and .539 respectively).

3. When Integrative Orientation was used as a moderator, i.e.- to classify Ss as High/Low on Orientation, the difference between the multiple R's of the two groups (.570 and .516 respectively) was not significant.

4. When Achievement Motivation was used as a moderator, i.e.- to classify Ss into High/Middle/Low one-third sub-groups the differences in multiple R's (.544, .631, and .482 respectively) did not reach significance at $P < .05$.

5. When both Integrative Orientation and Achievement Motivation were used as moderators, i.e.- to classify Ss according to various combinations of levels of Orientation and Motivation, no differences in multiple R's ranging from .432 to .643 reached significance at $P < .05$.

Perhaps at this point it might again be stated that the statistical tests applied to arrive at conclusions 3, 4, and 5 did not yield significant results at $P < .05$. Incidentally, those readers who do not disdain the "interocular traumatic test" or the "eye-balling approach" might want to do their own thinking regarding such a range in multiple R's.

6. For the comparisons of multiple R's for subgroups A-K inclusive to the multiple R for the total group (T) which included the subgroups (See Table 1, Chapter 1), a significant difference was found between the multiple R for the middle Motivation subgroup and the multiple R for the total group ($P < .05$). Using only the aptitude sub-tests as predictors, the multiple R for the middle Motivation subgroup ($R = .631$) was found to be significantly higher than the multiple R for the total group ($R = .539$). All other differences did not reach the .05 level of significance.

7. When multiple R's for subgroups G and H (R's of .570 and .516 respectively when Integrative Orientation used as a moderator) are each compared to the multiple R for the total group ($R = .540$ when Integrative Orientation included as a predictor in addition to the MLAT sub-tests) no significant difference was found. The Integrative Orientation variable was not found to improve prediction, whether used as a moderator or as a predictor.

8. When multiple R's for subgroups I, J and K (R's of .482, .631 and .544 respectively when Achievement Motivation is used as a moderator) are each compared to the multiple R for the total group ($R = .546$ when Achievement Motivation included as a predictor in addition to the MLAT subtests) no significant differences were found between the multiple R's for subgroups I and K (Low and High motivation groups respectively) and the multiple R for the total group. A significant difference was found, however, between the multiple R of subgroup J (Middle Motivation group) and the multiple R for the total group. The multiple R for the Middle Motivation subgroup (.631) was found to be significantly higher than the multiple R for the total group (.546). In other words, the motivation measure functioned better as a moderator than as a predictor for the Middle Motivation subgroup comparison to the total group ($P < .05$).

9. When multiple R's for subgroups A, B, C, D, E, and F (R's ranging from .432 to .643 when Integrative Orientation and Achievement Motivation joinly used as moderators) are each compared to the multiple R for the total group ($R = .547$ when Integrative Orientation and Achievement Motivation both included as predictors in addition to the aptitude subtests) all differences in multiple R's did not reach significance at $P < .05$. In other words, no improvement in prediction was found whether the orientation or motivation variables were used as moderators or as predictors.

10. No support for the interactive or multiplicative effects of aptitude, motivation, and orientation on performance in intensive French language learning was found although a trend to statistical significance for the A:B interaction (Aptitude:Motivation) was noted.

For the next two conclusions no implications for hypotheses testing are stated. Rather they represent what was intended as further support for some of the earlier stated findings.

11. The mean double cross-validation R's for each of the Low/Middle/High Achievement Motivation subgroups (R's of .41, .63 and .38 respectively when motivation used as a moderator) are compared to the mean double cross-validation R for the total group, (R of .44 when motivation included as a predictor in addition to aptitude and orientation variables). No significant differences were found for the Low and the High sub-group comparisons to the total group. For the Middle Motivation sub-group comparison however, a significant improvement in prediction was found in favor of the motivation measure used as a moderator ($P < .01$).

12. In the post-analyses, when the Integrative Orientation variable was dropped and a 2 X 3 least squares analysis of variance re-computed, the interaction effect of aptitude and motivation on performance in intensive French language learning was not found to be significant at $P < .05$ although trends to significance are again noted.

In general then, the findings of this research support the belief that a measure of motivation can function better as a moderator than as a predictor in selection research models. The findings strictly apply to the type of population presently undergoing training in the Canadian public service intensive French language training program. Because of the fact, however, that the population participating in the research is heterogeneous with regard to age, education, and socio-economic background,

and that no pre-selection had occured on relevant variables prior to training, it might well be that the findings have more generality than the writer is prepared to make at this time. As an example, the Canadian public Service population supplying the Ss in this research would appear to be quite comparable to the population Lawler[49] selected in a different context to test for interactive or multiplicative effects of ability and motivation on performance. A similar reference could be made to the US Peace Corps studies[50] in foreign language learning where similar predictor and criterion variables were used.

Again, the findings reported strictly apply to the specific moderator and multiple regression strategies utilized in this research. Some interesting extensions to this investigation could involve the testing of additional models with the same prediction and the same criterion measures. In addition, other variables such as age, sex, etc..., could be investigated as either potential moderators or predictors. As one example, Carroll[51] states that we are unlikely to

---

49 Lawler, Op. Cit., p. 153-163.

50 John B. Carroll, A Parametric Study of Language Training in the Peace Corps, Harvard University, Final Report for Contract PC-(W)-226, 1966, p. XIV-146.

51 John B. Carroll, and Stanley M. Sapon, Modern Language Aptitude Test MLAT Manual, 1959, Edition, The Psychological Corporation, New York, p. 23.

acquire research evidence in the near future regarding rates of learning a foreign language for individuals of different age groups. A reason cited is that it is not usual in the typical school situation to have individuals of various ages in the same learning situation. In this research the population is heterogeneous with regard to age.

Another possible extension might be to initiate a follow-up of Ss to determine if the findings hold using a retention/drop-out criterion. This would provide a re-test and extension of Datel et al's[52] finding that the Achievement Motivation Scales of the California Psychological Inventory differentiated between Ss who "remained in" an intensive foreign language program and those who "dropped-out" while a standard aptitude test battery did not.

There might also be some advantage in investigating sequential prediction strategies. It was found in this research, for example, that the correlations among the speaking, writing, listening and reading performance scores of the MLA-LB Achievement Test Battery were highly correlated. It might be as Carroll[53] suggests that at more advanced stages of language knowledge and use these correlations would decrease.

---

52 William E. Datel, Forrest D. Hall, and Charles P. Rufe, "Measurement of Achievement Motivation in Army Security Agency Foreign Language Candidates", in Educational and Psychological Measurement, Vol. XXV, No. 2, 1965, p. 539-545.

53 Carroll et al, Op. Cit., p. 23.

BIBLIOGRAPHY

Banas, P.M., An Investigation of Transituational Moderators, Unpublished Doctoral Dissertation, University of Minnesota, 1964.

This research provides a survey of the historical development and status of research on moderator variables up to 1964 and proposes a definition of a moderator which was accepted for this research.

Carroll, J.B., A Parametric Study of Language Training in the Peace Corps, Fiscal Report for Contract PC-(W)-226 with the Peace Corps, Harvard University Press, 1966, 146p.

This report presents data on foreign language learning in both the training and field-duty aspects of Peace Corps programs. From the results of this study decisions were made on some of the aptitude and achievement variables to be adopted in this research.

Cooley, W.W., and Lohnes, P.R., Multivariate Procedures for the Behavioral Sciences, New York, John Wiley and Sons, 1962, p. 35.

This text calls attention to certain reservations about the popularly used Stepwise Multiple Regression Procedure and was instrumental in ensuring that cross-validation was conducted because of the serious danger of capitalizing on chance.

Darlington, Richard B., "Multiple Regression in Psychological Research and Practice", in Psychological Bulletin, Vol. 69, No. 3, 1968, p. 161-182.

This theoretical article defends stepwise regression and extensions thereof, and shows that other selection procedures are inferior to the stepwise method.

Datel, William E., Hall, Forrest D., and Rufe, Charles F., "Measurement of Achievement Motivation in Army Security Agency Foreign Language Candidates", in Educational and Psychological Measurement, Vol. XXV, No. 2, 1965, p. 539-545.

The research evidence in this study was instrumental in the decision to adopt the Achievement Motivation scales of the California Psychological Inventory as a relevant measure of the motivation dimension in intensive language learning.

Fleishman, Edwin A., "A Relationship Between Incentive Motivation and Ability Level in Psychomotor Performance", in Journal of Experimental Psychology, Vol. 56, No. 1, 1958, p. 78-81.

The research evidence is presented in support for the validity of Vroom's interactive or multiplicative model.

French, E.G., "Effects of Interaction of Achievement Motivation and Intelligence on Problem - Solving Success", in American Psychologist, Vol. 12, 1957, p. 339-400.

This research with US Air Force airmen provides further support for the validity of Vroom's interactive model. The means on problem solving for the high motivation groups were separated according to intelligence level while those for the low motivation groups were not.

Galbraith, Jay, and Cummings, L.L., "An Empirical Investigation of the Motivational Determinants of Task Performance: Interactive Effects Between Instrumentality - Valence a and Motivation - Ability", in Organizational Behavior and Human Performance, 2, 1967, p. 237-257.

Vroom has suggested that performance can be thought of as a multiplicative function of motivation and ability. Results of this research are interpreted in terms of and support for Vroom's model.

Gardner, Robert C., and Lambert, Wallace E., "Motivational Variables in Second Language Acquisition", in Canadian Journal of Psychology, 1959, 13(4), p. 266-272.

Results of this research demonstrate that two factors, not one as previously presumed, are associated with second language achievement. The fact that Ss with an Integrative Orientation perform better than Ss with an Instrumental Orientation influenced the decision to use the integrative measure in this study.

Guion, Robert M., and Gottier, Richard F., "Validity of Personality Measures in Personnel Selection", in Personnel Psychology, Vol. 18, 1965, p. 135-164.

The theme of this article and Guion's suggestion that motivational variables may serve better in the role of moderator variables than as predictors led to the primary goal of this research.

Herzberg, Paul A. "Parameters of Cross-Validation" in Supplement to Psychometrika, 1969, p. 1-70.

This publication provided a comparison of different multiple regression methods in terms of cross-validation. It also points out the desirability of comparing prediction methods on a wide variety of data sets.

Hinrichs, John R. "Psychology of Men at Work", in Annual Review of Psychology, Vol. 21, 1970, p. 535.

This appraisal supports the need for comprehensive model testing with more heterogeneous populations and more reliable measures, and was to some extent influential in the decision to test for interactive or multiplicative effects of ability and motivation on performance in the context of second language learning.

Lawler, Edward E. III, "Ability as a Moderator of the Relationship Between Job Attitudes and Job Performance", in Personnel Psychology, Vol. 19, 1966, p. 153-164.

This research provides support for Vroom's model which suggested job performance to be some function of ability times motivation. There was also the observation that the population used by Lawler was quite similar in respect to age, salary level, position, and education to the population selected for this dissertation.

McCornack, Robert L., "A Comparison of Three Predictor Selection Techniques in Multiple Regression", in Psychometrika, Vol. 35, No. 2, June 1970, p. 257-271.

In the above research, an iterative multiple regression model was found to select predictors with slightly higher cross-validities than the popularly used stepwise method. With the controversy in the literature over the optimum method of multiple regression, this influenced the decision in this study to at least test two extreme models to determine the one that would hold-up better on cross-validation.

Owens, W.H., and Jewell, D.O., Personnel Selection", in Annual Review of Psychology, Vol. 20, 1969, p. 434-436.

The evaluation of the search for enhanced differential validities via the moderator variable approach is according to these authors a promising move in the right direction, although some of the problems and the form it will ultimately assume is only beginning to emerge.

Saunders, David R., "Moderator Variables In Prediction", in Educational and Psychological Measurement, Vol. 16, 1956, p. 209-222.

This report is important to this research in that it was Saunders who originally proposed the concept of a moderator model.

Zedeck, Sheldon, "Problems With the Use of Moderator Variables", in Psychological Bulletin, 1971, Vol. 76, No. 4, p. 295-310.

In this recent theoretical article some of the problems encountered with the use of moderator variables are elucidated. Different methods for comparing differential validities are presented and evaluated.

APPENDIX I

AGE FREQUENCY DISTRIBUTION FOR Ss IN INTENSIVE FRENCH LANGUAGE TRAINING PROGRAM

Age Frequency Distribution for Ss in Intensive French Language Training Program

Ages	Frequencies
60-61	2
58-59	4
56-57	3
54-55	9
52-53	17
50-51	23
48-49	26
46-47	21
44-45	22
42-43	22
40-41	21
38-39	17
36-37	14
34-35	18
32-33	16
30-31	14
28-29	11
26-27	18
24-25	13
22-23	9
N =	300

APPENDIX II

RAW SCORE DATA FOR PREDICTOR AND CRITERION VARIABLES USED IN THIS RESEARCH

APPENDIX II

Raw Score Data for Predictor and Criterion Variables Used in this Research

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
907	11	21	38	19	22	103	25	180
1202	35	23	16	10	3	95	22	90
604	16	12	9	8	8	76	21	51
606	17	22	7	16	5	85	17	34
610	18	19	7	17	10	84	22	58
702	6	19	18	29	3	97	28	22
805	39	27	17	17	15	89	15	208
806	26	17	17	26	7	104	17	226
309	32	25	15	22	16	80	18	104
110	8	30	9	29	10	94	17	202
105	27	25	23	29	14	83	21	158
2402	36	21	17	25	24	83	13	117
1303	34	24	27	34	14	76	12	40
1406	18	26	24	3	3	74	25	138
1410	20	25	18	12	5	73	26	200
3303	38	23	27	10	12	97	25	109
2806	41	26	22	31	17	102	15	188
2803	40	30	34	23	19	100	20	178
2802	38	21	17	32	20	94	22	177
2512	29	21	18	19	22	93	21	224
2511	30	17	26	21	19	72	21	198

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG ORIENT.	MLA (LB)
	1	11	111	1V	V			
2510	40	23	20	18	12	96	12	202
2808	28	23	20	27	9	75	22	183
3907	34	27	22	33	19	64	28	163
507	25	17	14	11	6	85	20	121
508	19	21	13	23	11	98	22	101
603	38	24	22	28	15	86	16	88
605	16	19	17	22	13	89	24	77
607	30	21	9	33	8	86	20	87
608	13	15	7	18	11	91	20	77
609	43	21	22	23	12	93	24	94
701	41	20	26	14	11	79	17	55
703	28	23	6	16	8	83	16	44
704	22	20	9	20	3	97	24	59
705	20	21	10	11	10	85	15	51
801	38	23	26	29	14	102	23	209
802	40	28	34	38	19	87	23	253
803	35	24	20	27	10	93	11	241
804	41	28	30	25	16	76	23	223
807	24	19	16	27	17	86	20	206
808	43	30	20	37	18	82	23	219
809	43	27	42	36	24	98	22	234

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
810	36	28	20	30	9	92	14	241
901	40	23	15	22	5	90	21	212
902	37	28	35	25	10	105	17	243
903	27	24	23	14	8	88	23	207
904	25	27	23	28	0	93	11	209
906	39	22	26	22	12	95	18	192
908	23	21	20	33	16	98	20	236
909	42	23	16	35	17	89	8	230
1002	41	29	27	35	15	73	19	209
1003	39	30	22	26	13	66	19	217
1004	35	22	11	26	13	98	18	161
1005	6	21	20	11	6	86	19	151
1006	40	28	4	26	12	76	21	208
1007	43	20	26	18	22	88	22	172
1008	34	20	33	24	19	95	18	185
1009	40	23	10	22	9	91	18	173
1101	36	23	13	15	1	69	19	197
1102	39	23	19	22	9	90	26	162
1103	34	25	23	20	9	88	17	197
1105	31	23	17	14	12	99	23	177
1106	39	19	29	17	13	82	10	151

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
1107	25	19	14	17	2	98	12	153
1108	35	24	26	33	20	85	19	165
1201	42	21	33	18	19	93	16	126
1203	10	23	10	21	7	92	23	126
1204	20	22	20	19	11	93	23	116
1205	30	22	24	18	10	93	21	126
1206	39	22	17	23	10	87	10	122
1301	36	18	10	16	21	58	23	79
1302	28	19	12	18	6	93	10	50
1304	30	25	18	20	7	100	19	47
1305	30	18	7	18	16	92	19	63
1306	37	28	7	19	14	89	12	85
1501	23	28	23	27	10	97	17	169
1502	41	23	21	29	14	91	8	200
1503	23	24	19	14	9	83	17	129
1504	33	22	31	24	17	97	13	90
1505	35	19	31	30	12	100	22	167
1506	33	18	20	21	8	88	21	194
1507	30	22	24	28	9	95	14	138
1508	35	29	24	14	7	61	11	158
1509	40	24	25	16	14	81	20	98

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
1603	16	18	3	8	7	84	18	42
1604	9	21	9	5	8	91	24	62
1605	3	15	10	12	1	87	23	49
1607	25	13	12	25	15	95	16	66
1608	14	25	34	14	5	81	25	46
1609	20	16	6	28	10	76	20	23
1702	17	18	7	14	18	96	18	58
1703	25	17	17	16	7	76	8	108
1707	36	19	40	26	14	94	10	75
1801	24	22	24	22	19	82	19	119
1802	35	25	23	11	15	98	16	118
1806	22	25	14	34	9	89	24	120
1807	37	22	19	16	22	60	20	164
1808	24	19	18	25	10	88	18	186
1809	18	20	22	17	17	85	11	167
1810	32	17	19	30	10	96	19	227
1902	39	23	24	23	19	98	22	157
1904	41	30	21	26	21	92	10	122
1905	31	17	8	23	10	99	18	180
1906	29	24	19	26	13	101	22	167
1908	29	25	22	21	7	95	18	102

APPENDIX II

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
1909	26	21	19	15	12	86	10	109
1910	38	22	16	24	10	72	25	84
2004	43	26	23	32	16	76	19	101
2005	16	12	15	32	18	90	17	83
2007	14	20	8	13	9	67	12	53
2008	21	14	29	14	15	97	21	105
2102	29	19	9	13	14	82	26	82
2103	25	23	8	20	11	94	13	167
2106	29	26	29	26	18	89	12	182
2107	39	27	16	19	4	84	18	127
2108	33	27	41	34	20	83	19	152
2109	19	19	16	29	19	89	15	131
2203	8	23	16	25	9	97	15	68
2210	36	18	13	21	8	95	14	89
2302	43	26	30	23	19	104	20	211
2306	43	24	24	34	24	84	25	230
2308	37	23	30	33	17	96	17	174
2309	35	25	24	36	12	97	24	231
2401	19	18	17	17	8	84	21	134
2403	33	16	18	20	8	103	23	101
2404	29	24	25	21	7	92	20	91

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
2405	24	18	28	9	13	79	14	72
2406	25	18	10	22	9	76	19	101
2407	33	21	12	26	13	97	11	135
2408	28	19	14	28	1	89	16	99
2409	5	25	21	9	10	90	17	78
2410	39	25	21	16	11	80	20	102
101	34	22	25	33	23	85	15	229
102	42	28	18	24	22	95	22	205
104	43	26	22	14	22	90	22	172
106	30	26	16	23	22	95	14	199
107	32	16	10	22	6	95	23	196
108	21	23	37	27	4	88	17	211
109	40	24	4	14	13	81	20	243
201	18	16	7	21	2	97	24	156
202	36	22	23	21	22	83	21	141
203	23	17	8	13	9	72	17	119
204	31	22	10	23	3	86	23	126
205	1	20	10	14	2	98	20	180
206	1	16	15	18	6	85	16	109
207	22	29	10	23	6	92	17	173
208	11	25	24	10	7	74	19	160

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
209	36	25	27	30	18	81	22	153
210	39	21	16	28	13	68	12	199
301	23	20	7	15	7	95	21	105
302	41	21	14	31	10	93	19	159
303	10	21	20	23	18	93	24	135
304	38	23	9	17	14	80	22	151
305	33	27	23	20	7	90	16	144
306	31	18	9	16	8	86	27	108
307	32	22	14	23	21	95	24	117
308	39	20	31	20	14	94	19	139
401	43	24	14	21	17	86	19	178
402	23	14	10	13	12	83	21	105
403	39	26	14	25	6	85	18	144
404	15	25	28	21	20	87	19	123
405	18	24	16	27	16	94	22	117
406	42	28	17	26	18	80	23	128
407	39	22	21	13	14	83	18	105
408	35	21	18	19	4	88	23	106
409	26	21	6	24	12	90	20	181
501	28	18	24	15	10	92	22	65
502	33	20	27	29	18	85	22	109

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
503	43	21	23	13	18	78	21	90
504	33	17	8	20	12	98	20	112
505	36	21	20	17	7	72	18	80
506	22	20	21	25	5	77	12	104
3805	42	22	20	15	17	74	21	42
3702	18	24	18	20	9	89	19	58
3605	36	22	16	14	11	78	25	92
3508	34	17	27	21	14	87	19	107
2807	43	27	28	28	14	105	20	191
1704	41	19	23	13	20	99	14	65
103	37	21	24	17	12	88	20	184
2502	43	30	32	17	24	96	19	229
2503	32	23	16	35	23	92	22	253
2504	32	27	22	32	14	102	27	259
2505	43	25	42	28	18	68	19	229
2506	32	24	22	25	17	86	19	226
2507	41	28	29	30	24	99	19	245
2508	37	24	33	37	24	93	19	252
2601	34	28	10	21	18	94	20	221
2602	19	26	28	22	8	97	18	208
2603	36	28	22	14	10	77	17	219

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
2605	42	30	15	23	18	81	21	223
2606	18	22	19	13	9	98	23	159
2607	10	26	29	30	7	56	16	206
2608	38	25	31	26	24	81	17	213
2701	30	14	14	13	9	91	24	225
2702	24	6	6	18	14	61	13	202
2703	41	25	11	24	15	91	24	203
2704	36	21	11	17	10	93	21	209
2705	41	26	28	26	15	77	25	197
2706	43	27	11	37	8	79	12	221
2707	25	26	33	29	14	90	27	211
2708	37	29	39	33	24	86	26	213
2801	28	25	26	25	16	92	19	213
2804	40	19	18	29	12	85	16	235
2805	16	26	24	19	13	92	21	207
2809	34	26	30	37	14	78	28	195
2901	10	24	24	32	13	101	25	192
2902	28	23	17	25	21	86	19	196
2903	40	22	20	33	9	86	18	217
2904	12	23	30	30	9	82	22	161
2905	13	21	18	26	8	75	21	163

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
2906	28	23	16	16	12	75	19	188
2907	40	24	15	28	20	94	21	173
2908	43	23	23	37	17	93	14	196
2909	43	29	34	19	23	102	20	167
2910	37	26	27	22	23	87	25	181
3001	35	18	9	23	16	66	23	173
3002	12	19	16	19	12	69	19	195
3003	24	18	11	31	21	94	15	164
3004	12	13	19	21	6	87	13	176
3005	27	27	6	22	11	80	22	138
3006	41	26	27	24	12	91	21	191
3007	26	27	15	15	4	94	21	150
3008	42	17	24	19	6	93	21	171
3102	37	21	24	29	11	103	15	132
3103	32	20	12	27	20	75	21	173
3104	4	29	12	19	13	81	17	146
3105	12	24	21	39	16	102	19	179
3106	40	26	27	20	17	92	16	163
3107	41	21	8	34	13	76	18	141
3108	18	16	22	10	22	89	24	147
3201	38	23	27	10	12	97	25	125

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
3202	25	24	8	21	14	60	10	148
3203	41	28	30	13	7	85	20	156
3204	41	21	14	24	1	81	18	133
3205	30	26	32	22	5	102	15	211
3206	33	20	8	18	7	90	27	125
3207	29	26	18	31	12	89	21	148
3208	34	17	22	12	13	64	23	96
3209	43	23	23	32	11	77	14	123
3210	30	24	12	16	5	68	16	162
3301	15	24	12	18	13	97	20	83
3302	24	23	11	15	15	93	16	108
3304	30	17	14	29	9	72	21	100
3305	34	19	7	8	7	84	22	89
3306	41	26	12	21	23	86	21	114
3307	30	21	20	23	20	84	12	144
3308	23	23	7	17	8	72	27	100
3309	43	21	7	25	20	79	16	107
3401	37	19	28	26	17	93	18	143
3402	31	20	13	19	17	76	19	87
3403	18	18	5	15	12	82	27	131
3404	32	17	10	14	8	76	16	97

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
3405	21	23	4	11	4	88	21	95
3406	25	15	6	14	6	86	24	87
3407	31	20	32	8	13	86	27	168
3501	38	19	12	14	9	85	23	66
3502	40	23	14	22	19	87	18	115
3503	8	20	1	8	6	80	21	77
3504	24	29	30	14	13	97	20	95
3505	0	21	25	18	7	92	14	111
3506	40	22	11	9	11	75	18	93
3507	28	25	24	13	14	72	24	88
3601	30	21	14	18	17	95	22	64
3607	26	26	16	20	9	95	19	50
3608	25	26	34	17	9	97	21	84
3701	24	12	5	7	4	87	18	47
3703	39	26	32	26	9	92	21	79
3704	24	14	3	9	7	89	23	39
3705	36	22	12	26	16	72	13	26
3801	40	23	19	18	13	75	20	59
3802	40	20	9	22	11	82	23	51
3803	17	17	8	9	7	74	21	48
3804	28	24	18	13	12	75	7	89

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
3902	43	29	45	30	24	84	19	193
3903	40	30	48	32	24	103	19	189
3905	36	23	30	32	15	80	21	154
3906	43	29	47	41	18	98	15	201
3908	41	20	10	19	15	85	23	164
1401	41	29	41	35	18	103	24	215
1402	34	23	30	25	3	87	20	179
1403	43	24	30	35	17	87	26	198
1404	40	22	11	23	11	90	20	142
1405	20	19	19	26	13	96	19	218
1407	37	26	23	31	13	100	25	197
1408	43	27	28	30	9	94	17	131
1409	32	28	16	27	22	76	12	161
1411	35	26	23	21	19	100	20	201
1412	38	20	18	20	14	92	21	202
1413	41	24	11	22	21	96	16	196
1414	43	25	25	31	24	93	19	222
1415	43	29	37	28	13	98	18	189
1416	31	25	24	16	13	83	21	126
1417	28	12	8	14	5	89	8	180
1418	43	22	20	23	16	94	18	219

Raw Score Data for Predictor and Criterion Variables Used in this Research (cont'd)

SUBJECT	MLAT					CPI	INTEG. ORIENT.	MLA (LB)
	1	11	111	1V	V			
1419	43	28	20	18	16	81	17	130
1420	43	27	31	24	24	89	25	193
3603	21	21	15	12	9	73	16	94
3604	43	23	24	16	20	88	21	73
3606	30	25	17	21	12	97	21	63
3602	35	28	42	23	20	51	23	61

N = 300

APPENDIX III

COPY OF SCALE USED IN THIS RESEARCH TO MEASURE Ss INTEGRATIVE ORIENTATION

Copy of Scale Used in this Research to Measure Ss Integrative Orientation

---

NAME AND INITIALS: .....

IDENTIFICATION NO. (or SIN) .....

CLASS OR GROUP LEVEL: .....

Listed below are four reasons frequently given by individuals for studying French. You are requested to read each reason carefully and indicate the extent (on a seven point scale) it is descriptive of your own reasons for studying French. The rating scale below each question has the following forms:

Not my feeling at all __: __: __: ✓: __: __: __: Definitely my feeling.
(If you were neutral about the question, you would insert a check mark (✓) in space 4 as shown above).

1. It will help me to understand better the French Canadian People and their way of life.

Not my feeling at all __: __: __: __: __: __: __: Definitely my feeling.

2. It will enable me to gain good friends more easily among French speaking people.

Not my feeling at all __: __: __: __: __: __: __: Definitely my feeling.

3. It should enable me to begin to think and behave as French Canadians do.

Not my feeling at all __: __: __: __: __: __: __: Definitely my feeling.

4. It will allow me to meet and converse with more and varied people.

Not my feeling at all __: __: __: __: __: __: __: Definitely my feeling.

---

* Permission granted by Dr. Lambert of McGill University to Reproduce and Use the Scale.

APPENDIX IV

RAW SCORE PERFORMANCE DATA USED IN TESTING VROOM'S MULTIPLICATIVE OR INTERACTIVE MODEL

Raw Score Data, Means, Variances, SD's for MLA-LB (Criterion Measure) for Ss Classified According to Various Combinations of Ability Level[a], Achievement Motivation[b], and Integrative Orientation[c].

1	2	3	4	5	6	7	8	9	10	11	12
a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.
H H H	H H L	H M H	H M L	H L H	H L L	L H H	L H L	L M H	L M L	L L H	L L L
215	201	179	178	61	193	150	218	194	180	51	160
125	196	203	229	59	141	83	202	62	109	48	89
109	189	253	182	173	221	90	118	120	131	77	97
177	132	213	226	195	229	156	180	91	58	104	148
224	139	156	107	42	199	180	102	207	63	173	146
117	75	181	235	153	152	112	66	125	34	126	202
157	189	73	213	141	119	105	58	66	109	105	206
197	131	79	196	102	158	117	108	77	173	82	104
201	229	172	196	98	55	126	50	206	111	46	101
180	169	149	143	223	40	64	135	212	84	79	72
221	136	114	115	163	161	105	226	126	99	200	108
95	174	109	208	198	26	135	22	126	176	88	42
84	159	230	230	158	107	171	227	147	186	131	197
236	211	203	197	197	123	209	153	106	167	100	94
209	90	211	241	223	213	59	50	181	78	89	93
178	219	191	122	90	219	101	47	95	51	96	87
231	185	198	123	128	101	192	208	87	141	138	162
205	188	193	88	84	151	63	161	168	173	163	133
167	199	253	163	164	209	177	209	225	85	161	195
173	179	172	144	208	217	116	164	49	47	92	188
167	243	184	217	219	117	159	167	142	87	151	105

1	2	3	4	5	6	7	8	9	10	11	12
a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.	a.b.c.
H H H	H H L	H M H	H M L	H L H	H L L	L H H	L H L	L M H	L M L	L L H	L L L
211	192	169	200	154	104	192	68	77		243	119
259	126		122		130	101	89	207		134	127
94	65		222					39		23	53
234	245		211					164		183	129
191	202		144					108		138	44
167	241		144					65		58	80
			165					121		51	
			252								
N:											
27	27	22	29	22	23	23	23	28	21	28	27
MEAN:											
178.67	174.22	176.59	179.75	146.95	147.17	128.83	131.65	128.32	111.52	111.79	121.52
VARIANCE:											
2219.85	2297.06	2248.88	3124.55	3067.95	3407.19	1893.97	4257.62	2940.43	2318.54	2662.31	2243.81
SD:											
47.12	47.93	47.42	55.90	55.39	58.37	43.52	65.25	54.23	48.15	51.60	47.37

LEGEND:

(a) For Ability Measure ≤(MLAT), Ss Split at Median into High (H) and Low (L) Groups.

(b) For Achievement Motivation Measure (CPI), Ss Split into High 1/3 (H), Middle 1/3 (M), and Low 1/3 (L) Groups.

(c) For Integrative Orientation Measure Ss Split at Median into High (H) and Low (L) Groups.

APPENDIX V

ABSTRACT OF

DOES EITHER A MODERATOR MODEL OR A MULTIPLICATIVE MODEL IMPROVE PREDICTION OVER THE USUAL MULTIPLE REGRESSION MODEL IN THE CONTEXT OF AN INTENSIVE FRENCH LANGUAGE TRAINING PROGRAM?

APPENDIX V

ABSTRACT OF

Does Either the Moderator Model Or the Multiplicative Model Improve Prediction Over the Usual Multiple Regression Model?

Prior to their participation in an intensive French Language training program in the Canadian Public Service language schools in Ottawa, 300 Ss at the lower and middle management levels were administered a battery of aptitude, motivation and orientation tests. At the end of three weeks of intensive language training the same Ss were administered a standardized battery of French Language achievement tests. The research was primarily designed to determine whether measures of motivation and orientation would function better as moderators than as predictors in the context of French Language learning. As a secondary aim the research was designed to test for interactive or multiplicative effects of ability, motivation, and orientation on performance. For a middle motivation subgroup a significant incremental validity was found ($P < .05$) in favor of the motivation variable used as a moderator. This improvement in prediction held up on cross-validation

---

1 Charles E. McInnis, doctoral thesis presented to the Faculty of Psychology of the University of Ottawa, Ottawa, Ontario, July 1970, xii-167p.

($P < .01$). For all other comparisons the differences did not reach statistical significance. The interactive effects of ability, motivation, and orientation on performance were not found to be significant at $P < .05$ although trends to significance are noted.