



27942551

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 00777 6804

**LIBRARY  
Michigan State  
University**

This is to certify that the

dissertation entitled

Cross Validity of Authentic- and  
Proxy-Criterion Regression Methods in the  
Selection of Veterinary School Applicants

presented by

Ivan Alexander Stuck

has been accepted towards fulfillment

of the requirements for

Doctor of Philosophy

degree in

Educational Measurement,  
Evaluation, and Research Design

*William A. Medrens*

Major professor

Date 12-5-1989

**PLACE IN RETURN BOX** to remove this checkout from your record.  
**TO AVOID FINES** return on or before date due.

DATE DUE	DATE DUE	DATE DUE
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

**MSU Is An Affirmative Action/Equal Opportunity Institution**

c:\circ\datedue.pm3-p.

**CROSS VALIDITY OF AUTHENTIC- AND PROXY-CRITERION REGRESSION  
METHODS IN THE SELECTION OF VETERINARY SCHOOL APPLICANTS**

**By**

**Ivan Alexander Stuck**

**A DISSERTATION**

**Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of**

**DOCTOR OF PHILOSOPHY**

**Department of Counseling, Educational Psychology, and  
Special Education**

**1989**



## **ABSTRACT**

### **CROSS VALIDITY OF AUTHENTIC- AND PROXY-CRITERION REGRESSION METHODS IN THE SELECTION OF VETERINARY SCHOOL APPLICANTS**

**By**

**Ivan A. Stuck**

Several advantages might be gained for admissions departments from the use of a proxy criterion in the development of a predictive multiple regression equation for selecting among candidate characteristics: (1) moderator error might be avoided by restricting prediction to only the immediate group of applicants, (2) predecessor data might not be necessary to develop a precise predictive equation, (3) novel criterions or predictors might be entered into the predictive equation for immediate use, (4) an interval-style multiple regression procedure might be used despite graduate level pass/fail grading, (5) the range restriction problem could be avoided, and (6) the data from all applicants could contribute to the reliability of the predictive equation. Admissions and performance data for five cohorts of veterinary applicants were used to compare four proxy criterion methods with the conventional multiple regression approach to the

development of a predictive selection equation. The authentic criterion was the graduate grade-point average, while the proxy criterion was an undergraduate pre-veterinary studies GPA. Predictors were undergraduate GPA, admissions test scores, employment ratings, and biographical and other data. Prediction factors were developed by selecting on college origin and performance level, by varying calibration sample sizes, and by restricting the use of intercorrelated predictors. T-tests and MANOVAs were used to evaluate mean differences in prediction error among the conditions. When prediction error (in ranks) was transformed to emphasize the error for cases near the cut score, no method differences were significant among any prediction methods and neither was there a year effect. It was evident also that no methods differed from prediction by UGPA alone. When transformed prediction error (actual) for a proxy-criterion method was observed across five years, a year effect appeared among proxy conditions. The significant effect for year, nevertheless, was attributable to exceptional interactions among prediction factors for two of the five annual cohorts.

Several improvements to proxy criterion calibration are suggested and the potential for proxy criterion use at other sites and for other graduate programs is discussed.

**TO THE STUDENTS WHO CHALLENGE CONVENTION**

## ACKNOWLEDGMENTS

I have deeply appreciated those persons and organizations who helped me in various ways to complete this dissertation. Former colleagues at the Ingham County Department of Social Services, St. Vincents Childrens Home, John Eicher, the College of Veterinary Medicine, Bob Simpson of B-O-C Powertrain, and Dr. Juan Olivarez of Grand Rapids Public Schools, provided a means of flexible employment during my doctoral study. Dr. Richard Houang, formerly with our department, was particularly helpful in guiding the development of my initial dissertation proposal. Dr. John Tasker and Pat Lowrie of the College of Veterinary Medicine provided the research environment from which my thesis arose, and further made available the data for my research, and the continuing use of their college facilities. Aleta Zamel provided invaluable assistance in illuminating several subtle features of the data set. My friend Dr. Jeff Mayer, helped to get me started in programming SPSS, and the consultants at the Computer Center assisted me often in the later data analyses. I thank my committee, Dr. Bill Mehrens for his tireless editing and suggestions, Dr. Steve Raudenbush for his statistical expertise, and Doctors Fred Ignatovich and Jim Haf for their insightful advice. I thank Rosalind Goodman for correcting my spelling and other shortcomings.

## TABLE OF CONTENTS

	Page
<b>LIST OF TABLES . . . . .</b>	<b>.vii</b>
<b>LIST OF FIGURES . . . . .</b>	<b>x</b>
 <b>Chapter</b>	
<b>I. Introduction . . . . .</b>	<b>1</b>
Need . . . . .	1
Purposes . . . . .	6
Research Hypotheses . . . . .	7
Rationale for Research Hypotheses . . . . .	8
Hypothesis A . . . . .	8
Hypothesis B . . . . .	8
Hypothesis C . . . . .	9
Hypothesis D . . . . .	10
Hypothesis E . . . . .	10
Overview . . . . .	11
 <b>II. LITERATURE REVIEW . . . . .</b>	 <b>13</b>
Part 1: Substantive Review . . . . .	13
Summary . . . . .	21
Part 2: Theoretical Review . . . . .	21
Sampling . . . . .	21
Measurement . . . . .	24
The Reliability of Validity Coefficients . . . . .	25
Multiple Regression . . . . .	31
Multivariate and Univariate Analysis of Variance and the T-Test . . . . .	34
 <b>III. PROPOSED THEORY . . . . .</b>	 <b>38</b>
General-criterion (GC) Regression Approach . . . . .	39
Using Multiple Regression to Shrink Error . . . . .	44
A.The case where the school variable is absent . . . . .	45
B.The case where the sample has a predominate rule . . . . .	45
C.The case where rules vary within the school . . . . .	46
D.Where prediction is improved for extreme cases . . . . .	46
E.The case where a proxy criterion is used . . . . .	47
Proxy-criterion Alternatives to Conventional Prediction . . . . .	48
Local-proxy (LP) . . . . .	49

Chapter		Page
	General-proxy (GP) . . . . .	54
	General-criterion-mixed or General-mixed (GM) . . . . .	57
	Summary . . . . .	65
IV.	DESIGN . . . . .	66
	Population Sample . . . . .	66
	Predictors . . . . .	67
	Criteria . . . . .	69
	Analyses . . . . .	70
	Test of Methods (MANOVA <sub>TOM</sub> ) . . . . .	70
	Sample . . . . .	70
	Conditions . . . . .	70
	Dependent Variable . . . . .	72
	Satisfaction of Assumptions . . . . .	73
	Application to Hypotheses . . . . .	73
	Test of Factor and Year Effects (MANOVA <sub>TOF</sub> ) . . . . .	73
	Sample . . . . .	73
	Conditions . . . . .	74
	Dependent Variable . . . . .	75
	Application to Hypotheses . . . . .	75
V.	RESULTS . . . . .	75
	Test of Methods across Two Years . . . . .	77
	Contrasts Between UGPA and Local-proxy	
	Estimation . . . . .	77
	Comparisons Against the General-criterion	
	Method . . . . .	81
	Test of Years and Other Factors under LP	
	(Local-proxy) . . . . .	90
	Independent Variable Contrasts . . . . .	90
	Relative Validity of Non-MSU and MSU UGPAs . . . . .	101
VI.	DISCUSSION . . . . .	103
	High UGPA Validity . . . . .	105
	Potential Usefulness of Proxy Methods . . . . .	106
	Potential Improvements to Estimation with a	
	Proxy Criterion . . . . .	107
	How Intercorrelation May Remain Benign . . . . .	111
	Extreme Outcomes in Figures 5 and 6 . . . . .	112
	Cautions Regarding Study Realism . . . . .	115
	Why R <sup>2</sup> Wasn't Used as a Measure of Validity . . . . .	120
	Regression Equations Have Superfluous	
	Predictors . . . . .	122
	Practical Implications of the Study . . . . .	122
VII.	SUMMARY . . . . .	132

Chapter	Page
APPENDICES . . . . .	140
Substitute Merit Values . . . . .	141
Error Weight by GPA Plot . . . . .	143
Correlations: LEWAR-transformed Error by Predictors and GGPA . . . . .	146
REFERENCES . . . . .	153

## LIST OF TABLES

Table	Page
01 Cross-institutional disparities in predictors, predictor weights, and prediction validities . . . . .	14
02 Comparative validity: Ordinary Least Squares, Local-proxy, General-criterion, and college GPA . . . . .	17
03 Regression bias factors based on Richards (1982) . . . . .	28
04 Channels by which error may affect calibration methods . . . . .	62
05 Potential systematic error for each calibration method . . . . .	63
06 Potential random error for each of four calibration methods . . . . .	63
07 Potential advantages for each of four calibration methods . . . . .	64
08 Disadvantages of each of four calibration Methods . . . . .	64
09 Optimal conditions for use of each calibration method . . . . .	65
10 The study sample: Five cohorts of veterinary applicants . . . . .	67
11 Predictors used in levels of the intercorrelation factor . . . . .	69
12 Confirmation of expected outcomes . . . . .	77
13 Contrasts of UGPA prediction against LP prediction . . . . .	81
14 Method efficiency in logs of absolute rank-error . . . . .	83



Table		Page
15	Estimate of average contrast between general-criterion and general-mixed transformed prediction error means . . . . .	84
16	Test of main effect for methods (MANOVA <sub>TOM</sub> ) .	84
17	Test of main effect for years (MANOVA <sub>TOM</sub> ) . .	85
18	Cross-validities for prediction methods . . .	86
19	Test of main effects for years (MANOVA <sub>TOP</sub> ) . .	92
20	T-test of mean error between MSU and non-MSU applicants . . . . .	101
21	A chart for identifying ideal method prediction factors . . . . .	109
22	Method betas under optimal prediction conditions . . . . .	125
23	Predictor names and their definitions . . . .	126
24	LP betas by year: full sample . . . . .	127
25	LP betas by year: 75% sample . . . . .	128
AI	Values substituted for missing merit values .	142
AIII	LEWAR transformed error by predictors and GGPA correlations . . . . .	147

## LIST OF FIGURES

Figure	Page
1      Relative levels of transformed prediction error among GGPA estimation methods (1984) . . .	88
2      Relative levels of transformed prediction error among GGPA estimation methods (1985) . . .	89
3      Prediction error under local-proxy prediction where the proxy-criterion has been restricted . . . . .	93
4      When measured by transformed prediction error, local-proxy prediction appears little affected by level of intercorrelation . . . . .	94
5      When measured by transformed prediction error, restriction of the proxy- criterion range only appears to affect prediction for the 1984 cohort . . . . .	96
6      When measured by transformed prediction error, source of UGPA only appears to affect prediction for the 1984 and 1985 cohorts . . . .	97
7      When measured by transformed prediction error, local-proxy prediction appears affected by level of intercorrelation for the 1984 cohort .	98
8      When measured by transformed prediction error, sample size appears to have little effect on prediction . . . . .	99
AI Ia   LEWAR/LEWA weights plotted by GGPA rank (n= 88) . . . . .	144
AI Ib   LEWAR/LEWA weights plotted by GGPA (n= 88) . . . . .	145

## CHAPTER I

### INTRODUCTION

#### Need

The efficiency of any process and the quality of its output improves as the selection of its input becomes more and more purposeful. Concurrent refinement of both (1) the criteria identifying output quality and (2) selection of input is additionally beneficial as an approach to quality development.

When the process is graduate level education and the outcomes are licensed practitioners, the selection process is administered by a college department of admissions, while the criteria are largely determined by both (1) a host of university instructors, and by (2) professional examination boards. In such a context, selection and criteria tend to become estranged. However, because the refinement of selection requires that knowledge of the identity and weights of valid predictors (which reside in the admissions department data) be linked to quality criteria (which remain under the domain of the student records authority) a cross- departmental flow of information is desirable. For many years the logistical difficulty of bridging the departmental offices was an authentic barrier to their reciprocation.

Today, however, inexpensive and adequate technology exists for the necessary data storage, integration, and analysis to allow admissions policy to be shaped by data on student performance. An equation which weights and combines admissions data (predictors) to estimate a performance outcome (or criterion) becomes a useful link between pre-program credentials (admissions data) and program performance (student records). Such an equation is known as a linear model. Perhaps the most useful tool in developing a linear model, (or "selection formula" for present purposes) is the multiple regression procedure. Multiple regression allows the use of past experience to inform present decisions. Given a criterion of quality (e.g. graduate school grades) and a set of application scores, the procedure can select the most predictive variables and weight them to maximize prediction of the criterion. This linear model which was optimal as a selection formula for the original data set may still remain useful for predicting the future grades of present students.

Unfortunately, research in this area has failed to demonstrate consistent outcomes. Typically, selection formulas will differ by location or by year of data studied (see Niedzwiedz & Friedman, 1976). For the most part, discrepancies are unsurprising, due to the limited sampling and sample sizes involved (single graduating classes of less than 100 students are typically used). Often, reports

mention only a limited number of predictors which were considered significant, leaving the reader to guess what additional predictors may or may not have been tried (see Niedzwiedz & Friedman, 1976; Hart, Payne, & Lewis, 1981; Markert, 1983; and Jones & Thomae-Forgues, 1984).

In the past, multiple regression research would demand substantial resources. Collection of admissions data would require increasing administrative costs and organization as the set of predictive variables was expanded. Additionally, prior to the advent of the computerized office, many hours of clerical labor were required for the transfer of both admissions and student performance data from office documents to a usable medium for data analysis. In addition to these obstacles, selection formulas obtained from one year could be notoriously unreliable for predicting performance for a subsequent year. The resource drain projected for a multiple-year regression study was considerable, and few admissions officers could be confident that the advantage would compensate for the loss entailed.

Today, fortunately, many of these previous costs have diminished due to the advent of the microcomputer. The evolution of methodological innovations may also provide cost effective improvements in the use of available data. One such potential multiple regression innovation may be the use of proxy-criteria (variables which measure a set of factors similar to those measured by the ideal criterion;

e.g. undergraduate grade-point average [UGPA] may serve as a proxy for graduate grade-point average [GGPA]).

The implementation of proxy criteria in multiple regression calibration studies may provide an alternate means of estimating a selection formula and extend the use of multiple regression selection. If the error between a proxy and the real criterion is of less importance than the error associated with the potential confounding due to years, location, or other factors, a selection formula calibrated with the use of a proxy criterion may be preferable for decision making.

This author found no evidence that a proxy approach had been studied prior to the author's own pilot study (Stuck, 1986). In that case, the proxy criterion (prerequisite veterinary GPA) allowed substantially higher predictive validity than a selection formula that was calibrated on one year's data from one location. In addition to using a proxy criterion with a sample limited to a single year and location (referred to as "local-proxy method" or "LP"), the proxy may substitute for the criterion in a multiple year and/or multiple location sample calibration (the "general-proxy method" or "GP") or the proxy criterion may substitute conditionally-- only where real criterion data are absent (the "general-criterion and proxy method" or "GM") for multiple year and/or multiple location samples.

To summarize, proxy criterion estimation could use one of the following forms:

- (LP) single site, single year proxy criterion multiple regression estimates of betas (predictor weights)
- (GP) multiple site and/or multiple year proxy criterion estimates of betas
- (GM) identical to GP except that a proxy substitutes for the criterion only where the criterion measure is absent from the case.

Besides allowing the use of additional cases to increase the sample size, use of an adequate proxy criterion in the LP, GP, and GM methods would allow the inclusion of the normal range of applicant ability, thus eliminating the need to correct subsequent weightings for restriction of range. For the LP method (using solitary year and location data), additional variables can be added to the set of predictors in any year. This could allow, for instance, the use of two optional admission tests (e.g. GRE: Graduate Record Exam, and MCAT: Medical College Admissions Test) with confidence that they would both be appropriately weighted in the selection formula. The LP method would also be useful where the necessary data from previous years are unavailable.

For neither the conventional calibration approach (GC), nor for the proxy-inclusive approaches (LP, GP, GM) is there literature addressing the nature of admissions prediction error. It is likely that such knowledge could be useful in several ways: (1) if error in predicted scores

is random, then the quest for additional predictors might be ill- advised, (2) knowledge of relative levels of random error among methods would be of value for the analyst in choosing a calibration approach, and (3) where moderated prediction error is evident, evidence of its characteristics may assist to control or to reduce such error by the introduction of new variables or transformations.

Of particular concern is the nature of variation by years and by locations. Should prediction error appear to be moderated by these factors, then the LP approach may be advantageous due to its year-specific calibration and its option of selecting a different set of predictors. If the error appears to be both random and unilevel (identically distributed) across factors, then the LP approach may provide no advantage over the conventional calibration.

### Purposes

The purposes of this research are to (1) create from veterinary school admissions data these four selection formulas: LP= a single year and location formula using a proxy criterion, GC= a conventional generalized formula using an authentic criterion, GP= a generalized formula using a proxy criterion, GM= a generalized formula substituting a proxy criterion only for cases which lack an authentic criterion; (2) to compare them in terms of



predictiveness on new applicant cases, and (3) to examine the nature of their prediction error.

### Research Hypotheses

This research is designed to test the following list of propositions (which precedes a subsequent commentary):

Hypothesis A. For students falling within a cut score zone, the local-proxy (LP) formula will be more predictive of graduate GPA (GGPA) than will undergraduate GPA (UGPA).

Hypothesis B. For students falling within a cut score zone, the general-mixed (GM) formula will be more predictive of graduate GPA (GGPA) than will the conventional prediction model (the general-criterion formula, GC) as corrected for range restriction.

Hypothesis C. With prediction error as the dependent variable, and with variation controlled with respect to years, methods, sample-size, academic origin, and intercorrelation of predictors, prediction differences among years and methods will be obtained.

Hypothesis D. With prediction error as the dependent variable, with methods limited to the local-proxy (LP) approach, and with variation controlled with respect to years, sample-size, academic origin, intercorrelation of predictors, and level of ability, prediction will vacillate across years.

Hypothesis E. Non-MSU undergraduates will be associated with greater prediction error.

## **Rationale for Research Hypotheses**

**Hypothesis A:** It is widely held that a previous grade-point-average (UGPA, for undergraduate study) is the best single predictor of future GPA (Mehrens and Lehmann, 1984). Therefore, given predictor cases limited to a single year and location, the UGPA would be expected to be the most reliable predictor of GGPA (graduate academic performance). Any proposed alternative (such as a formula formulated from the LP regression approach), therefore, must be able to outperform UGPA. Hence, Hypothesis A is to be evaluated by the relative validity of LP prediction against UGPA prediction. In addition, because predictive precision only matters where it may alter the conventional outcome, validity difference (between methods) must be granted more importance as it falls within the cut-score zone (the lower bound of the veterinary doctor achievement distribution). In this instance the weighting is done by a non-linear transformation of the errors of prediction which results in greater importance for the prediction errors for marginal veterinary students.

**Hypothesis B:** Where data extends across years or locations, the conventional GC (general-criterion) regression-formula validity is the standard (where corrected for restriction of range). Hence, for the proxy criterion alternative to prove itself useful, Hypothesis B requires that the GM (general-mixed) regression-formula validity must predict cut score proximity cases with less

error than does the conventional GC approach (corrected for range restriction).

Hypothesis C: Prediction by regression equation may vary according to specific factors (such as size or selection of the sample used in estimation of the regression equation). To determine the relative importance of such factors, it is necessary to control the influence of each factor. Control of prediction factors can be achieved by the deliberate selective sampling of calibration cases to bias selection formulas in a controlled manner (to deliberately exaggerate error effects), or by other deliberate means. Measurement of this bias is possible by applying the biased formulas to new data and by estimating the prediction error (between predicted and actual criterion scores). By entering these condition-specific prediction error values into a repeated-measures MANOVA procedure, the statistical and relative importance of these factors may be evaluated. Where such a MANOVA procedure is controlled for year and/or method moderators, the emergence of effects for the LP (local-proxy) and GP (general-proxy) methods should correspond with concurrent effects for years; if the dependent variable is not moderated, then the practice of cross-year and cross- location generalization will be unimpaired, and, hence, the conventional GC (general-criterion) approach may be preferable for use. Assuming that calibration factors

will not exhibit random influence, MANOVA effects for years and for methods are expected.

Hypothesis D: The following is consistent with the case of prediction parameters which vary across time: with a MANOVA procedure (1) limited to the local-proxy calibration method, and (2) controlled for year, sample-size, academic origin, redundancy of predictors, and level of achievement, an effect for years is expected. (Under the LP [local-proxy] approach, a larger sample-size and achievement range is possible, thus allowing the inclusion of an additional variable, achievement level, into the study). The likelihood of finding year effects is enhanced (1) due to the larger number of years which may be included, and (2) due to the additional variance available with the use of prediction error reported in interval scale. A moderator effect (such as a year effect) would suggest that (1) additional predictors are required in the regression equation or that (2) blocking is required on years. Blocking (e.g. local-proxy calibration) is a less precise means of control (than the addition of missing predictors), and therefore could be expected to only partly account for variance caused by changes in predictor validities.

Hypothesis E: The importance of the variation in UGPA standards across institutions may be confirmed by obtaining a significant difference in mean absolute error between subgroups which differ on UGPA origin (MSU vs non-MSU).

Another indicator is the contrast between calibration conditions which differ only on the N factor, but this would be a weaker test (M= MSU, N= all).

## Overview

Chapter II will present a two-part literature review: Part 1 will review the use of multiple regression in selection for health science and graduate school admissions, and Part 2 will review theoretical issues underlying the methods used in this study.

Chapter III presents the theory being examined by the present study. The relative effects of proxy-criterion use are hypothesized for three multiple regression approaches (LP, GP, and GM) in relation to the conventional multiple regression approach (GC). Also the potential for a systematic (vs. random) nature for prediction error is discussed.

Chapter IV outlines the designs for the two main analyses (repeated-measures MANOVAs) conducted within the research study: (1) the test of methods (TOM), a validity test comparing four multiple regression methods (GC, LP, GP, and GM) with respect to prediction error as the dependent variable; the methods each being controlled on three prediction factors (sample size, source of UGPA, and intercorrelation of predictors), and (2) the test of

factors (TOF), a validity test comparing prediction conditions and controlling an additional prediction factor--past academic performance in pre-veterinary courses, while holding methods constant (using the local-proxy, or LP method).

Chapter V presents the results of the study. From the methods test MANOVA (MANOVA<sub>TOF</sub>), the plausibility of five hypotheses will be judged: Hypothesis A, proposing the greater predictiveness of the LP model relative to that of the UGPA for marginal students; Hypothesis B, proposing the greater predictiveness of the GM model (that conditionally substitutes a proxy criterion value where the authentic criterion value is lacking) relative to that of the conventional multiple regression model (GC); Hypothesis C, proposing effects for years and methods; and Hypothesis E, proposing a greater association of prediction error for students claiming a non-MSU UGPA. In addition, the rank-order of method effectiveness relative to prediction-error will be observed. From the test of factors and years (MANOVA<sub>TOF</sub>), Hypothesis D will be tested (again) to confirm the existence of year effects.

Chapter VI offers a discussion of the findings.

Chapter VII offers a summary of the research.

## CHAPTER II

### LITERATURE REVIEW

#### Part 1: Substantive Review

Health Sciences candidate selection (including selection for human, veterinary, and dental medicine) provides an ideal domain for the study of academic selection because (1) the demand for medical education remains fairly consistent, and (2) medical education tends to remain uniform over time. Surprisingly, there have been few multiple regression studies of academic selection in this area, and none that this author has seen report any efforts to validate selection formulas longitudinally. Niedzwiedz & Friedman (1976) did study academic selection across schools, however. Table 1 shows disparity in the magnitude of correlations between predictors and the four-year veterinary school grades (ranging from  $r =$  non-significant to  $r = .55$ ). More important are the differences among sets of predictors. Assuming that similar scores and ratings are available to each institution for the evaluation of applicants, and assuming that the most predictive variables were reported, prediction appears to be inconsistent across schools. Additional studies (Hart, Payne, & Lewis, 1981; Markert, 1983; and Jones & Thomae-Forgues, 1984) found comparable

Table 1

Cross-institutional disparities in predictors, predictor weights, and prediction validities

Schools	Predictors	Criterion= GPA		
		r:Year1	r:Year2	r:Year4
<hr/>				
	[Niedzwiedz and Friedman, 1976]			
A	Physics GPA			.31
	Physics hours			
B	Chemistry GPA			.50
	Extra-Curricular Rating			
	VAT Total Score			
C	Science GPA			.55
	Academic Rating			
	VAT Science Score			
D	(not reported)			NS
 [Hart,Payne, and Lewis, 1981]				
E	College Science	.40	(w/biochem. mem.)	
E	College Science	.43	(w/biochem.intp.)	
E	College Science	.43	(w/biochm.p.lrn.)	
E	College Science	.39	(physiology)	
 [Markert, 1983]				
F	College GPA	.39		
	MCAT			
 [Jones and Thomae-Forgues, (1984)]				
S=25	College GPA	.41		
S=22	College GPA		.37	
S=25	MCAT	.41	.37	
<hr/>				
S=number of schools in study				



correlation magnitudes (all near  $r = .40$ ), but nevertheless, failed to demonstrate a reliable set of predictors of medical school performance.

Using Class of 1985 data as the regression formula calibration sample (to predict veterinary school performance), and attempting to validate the formula on 1986 and 1987 cohorts, the author obtained estimates correlating 0.49 with Vet School GGPA for 1986, but for 1987 the correlation dropped to a validity of 0.20. For 1987, however, there was a single variable by GGPA correlation as high as 0.41. Clearly, predictors selected for one class via multiple regression procedures may appear to be unreliable across subsequent classes. As is evident under these conditions, the regression formula may appear sufficiently unreliable that admission directors will feel justified in imposing subjective hunches or even prejudice into their selection processes, subsequently resulting in yet weaker and more prejudiced selection formulas.

Some efforts have been made to correct for error which contributes to unreliable selection formulas. In particular, attention has been directed towards error that occurs across locations. We know that considerable variation in academic standards exists from college to college. There are also many opportunities for deliberate and accidental transcription errors in the assessment of applicant credentials. Clapp and Reid (1976) improved prediction of medical student performance by weighting UGPA

by an index of undergraduate admissions standards. Linn (1966) reviewed research attempting to re-scale multi-standard applicant high school GPAs to a single, standard, scale (HSGPA). Although some prediction gains were observed for zero-order HSGPA X UGPA correlations by the use of a specific-school- adjusted HSGPA ( $HSGPA_s$ ), adjustments had no effect on multiple correlation coefficients where admissions test data were among the predictors. In all cases, where prediction gains were obtained for the validation sample, the prediction advantage shrunk substantially upon crossvalidity testing.

The use of the proxy criterion should increase sample size, at the expense of criterion precision; this may be preferable to accumulating potential year or location error from using additional years and/or locations as a means of increasing sample size. Wilson (1982) used UGPA as a proxy criterion in estimating the validity of GRE (graduate admissions test) scores. For chemistry majors (the reported major most relevant to the health sciences), he observed a correlation between UGPA and first year graduate GPA of 0.30 (pooled data for years 1974, 1975, and 1978;  $n=574$ ).

Stuck (1986) observed that the use of a proxy criterion (a pre-veterinary UGPA) might provide better estimation of veterinary school GGPA than may the use of a conventional multiple regression approach, because it can control for the potentially confounding effects of year and

location. This approach to developing a selection formula is referred to as the LP (local-proxy) approach. An LP selection procedure was carried out retrospectively for a set of Veterinary School applicants. Table 2 compares LP (local-proxy) prediction results with outcomes from a GC (general-criterion) model and the (optimal) ordinary least squares (OLS) correlation. The OLS equation predicts at  $R_{OLS} = .66$ . Because it is an original calibration, its predictors are uniquely selected and their weights are uniquely computed to minimize the squared error for that

Table 2

Comparative validity: Ordinary Least Squares, Local-proxy, General-criterion, and college GPA

RUN	APPLIC COHORT	CALIBRATION CRITERION	PREDICTORS	WEIGHTS	r (SCORE X GGPA)
OLS	1984	GGPA	1984	1984	.66
LP <sub>PVS</sub>	1984	PVUGPA	UGPA+VARS	1984	.58
LP <sub>CUM</sub>	1984	UGPA	PVUGPA+VARS	1984	.51
GC <sub>81</sub>	1984	GGPA	1981	1981	.20
UGPA	1984	-	UGPA	100%	.20

OLS= optimal equation validity for data set  
 LP= Local-Proxy prediction method  
 LP<sub>PVS</sub>= LP approach using PVUGPA (veterinary prerequisite UGPA) as a proxy-criterion  
 LP<sub>CUM</sub>= LP approach using UGPA as a proxy-criterion  
 GC= General Criterion (conventional) prediction method  
 GC<sub>81</sub>= GC approach applying 1981 regression equation to 1984 data

particular sample.  $GC_{81}$  uses a selection formula calibrated with a GC approach (having fixed predictors and fixed weights) computed from 1981 data which predicts at  $r = .20$  when applied to the 1984 data. The correlation for UGPA, which predicts with UGPA alone, is the same as that for  $GC_{81}$ , at 0.20. The LP formulas  $LP_{pvs}$  and  $LP_{cum}$ , predicting at 0.58 and 0.51, provide a better level of prediction.

The potential advantage of an LP (local-proxy) approach lies in its avoidance of moderator error from uncontrolled year, location, and other confounding effects. Where data is sampled in a non-random fashion--as is the case with admissions data--the presence of such effects must be expected unless there is substantial evidence to the contrary. Of course, year and location effects are, more precisely, artifacts of the changes in selection as it varies across years or locations. The importance of such moderation has been suggested by previous research. Gender, ethnicity, socioeconomic status, personality, sites, years, and high school rank are variables which have been found to moderate prediction coefficients. Doolittle and Cleary (1987) found that women do worse on math items. Hogrebe, Ervin, Dwinell, and Newmann (1983) reported differential validity among performance prediction models for gender for white (but not for black) ethnic subgroups. McCornack (1983) found white-ethnic subgroup differences for blacks, and Asians. Goldman and Hewitt (1976) found

that minority performance predictiveness differed even after controlling for specific program category. Wright and Bean (1974) found socioeconomic status to moderate prediction for a sample of white urban male college students. Neiner and Owens (1985) observed an association between vocational choice and personality factors, whereas, Gough and Lanning (1986) obtained male and female cross-validity coefficients of  $r = .38$  and  $r = .36$  with the California Psychological Inventory in predicting academic performance. Hakstian and Woolsey (1985), in turn, found validity coefficients for males and females of  $r = .39$  and  $r = .37$  with the California Aptitude Battery in predicting an introductory psychology course grade. Outside the health sciences area, Linn, Harnisch and Dunbar (1981) observed differences in LSAT validity for sites and years, additionally concluding that one cause appeared to be variation in grading as opposed to variable aptitudes. Goldman and Hewitt (1975) likewise found evidence of grading variability. Particularly, they observed an adaptation of grading standards relative to the ability range of the lower two-thirds of the class. Humphreys and Taber (1973) concluded from their postdictive studies that variation in grading standards best explained non-linear semester grade by GRE relationships. In part, this may be explained by differential attrition from academic disciplines (Loeb & Bowers, 1973) due to, in turn,

differential cross-discipline grading (Thornell & McCoy, 1985).

More recently, Elliot and Strenta (1988) were able to improve the validity of UGPA for prediction by adjusting scores according to departmental standards, thus reducing prediction bias for race and gender groups. McCornack and McLeod (1988) also reduced gender-related prediction bias by controlling for subject matter. Although Sawyer and Maxey (1979) found stable prediction over a four-year span in the prediction of UGPA from ACT (American College Testing) scores, Sawyer (1986) later found UGPA variation a major source of prediction bias, accompanied by the lesser sources of age, gender, and race. Wood and Langerin (1972) found that high school rank moderated prediction for high ability students.

It is recognized that where differential validity is inferred from discrepant correlation coefficients, the cause may often be artifactual due to (1) sampling error, (2) measurement error, or (3) the variability of the sample studied relative to the variability of the sample to which the equation is to be applied (commonly called the restriction-of-range problem, see Mehrens and Lehmann, 1984). Thus some of the preceding findings must be interpreted with caution, due to the uncertainty regarding control of artifactual effects.

## Summary

Moderation of selection formula validity by sites finds some confirming evidence in the health science education literature. Moderation across years, however, is more difficult to evaluate through literature review due to a dearth of longitudinal study of regression equation validity. The author's longitudinal study of the generalizability of a single-year, single-site, equation over years, found the validity to be poor. For one year, the use of a portion of the UGPA as a proxy variable allowed the calibration of a more valid regression equation for selection. Such an outcome may have been possible because of the presence of moderating factors associated with years. The literature reporting moderating factors is quite extensive. If use of proxy criterion regression avoids moderation by years, it nevertheless remains somewhat less valid.

## Part 2: Theoretical Review

### Sampling

Sampling theory provides justification for drawing inferences from samples under certain conditions. Suppose that a population exists who share some independently acquired mutual attributes and characteristics but who differ on other attributes and characteristics. If samples are drawn in large enough numbers and in a random manner, we are confident that: (1) randomly sampled, independently

acquired characteristics of the sample can be inferred to the population as a whole, and (2) randomly sampled, independently acquired characteristics of the sample can be inferred to any other large random sample of that particular population (these follow from the central limit theorem, see Huntsberger and Billingsley, 1973, pp. 131-134).

Ross (1988) cites Kish for classifying samples as (1) experimental, (2) survey, or (3) investigative, based largely on the quality of the sampling. An experiment provides deliberate treatment with control of extraneous variables by randomization or other means. A survey selects randomly from a defined population in which each member has a specific probability of being studied. In the investigation, however, control is the least. Sampling is by convenience with neither randomization nor probability sampling. The study of admissions data falls under this latter category. Where cases are not sampled randomly, but are selected according to their value on a particular variable (let's say "selected on IQ" [scholastic aptitude]), observed correlations between that selected variable (IQ) and another (say, academic performance) may be lower than would have been the case if the sample had been sampled randomly (an additional instance of the restriction-of-range problem).

If therefore, selection is on the dependent variable, or if the regression uses standardized variables,



regression /validity coefficients for selected data will be artificially low (Richards, 1982). This is the case in sampling to produce a multiple regression selection formula, wherein only selected applicant cases include a GGPA/ criterion. Because the validity coefficient reflects the proportion of true variance to error variance:

$$r_{xy} = s^2_T / (s^2_O = s^2_T + s^2_E),$$

a reduction in true variance resulting from selection of a restricted variable range leaves the error variance intact thus reducing the proportion of true variance to error variance. This is an instance of artifactual error, because the proportion of error is inflated due to the improper sampling procedures used.

Other error is due to the vagaries of the sampling process. Nuisance variables (Kirk, 1982), confounding variables, and moderator variables (Allen & Yen, 1979) are common labels referring to another factor which may reduce prediction validity during the sample selection phase of an investigation. Inasmuch as all members of a population will not equally share access to, nor interest in, graduate admissions; certain population traits and characteristics may be over-represented in a non-random sample of applicants. When such unspecified and uncontrolled-for variables affect performance on the dependent variable, an additional source of error is imposed on the investigation.

Two hazards accompanying the use of non-randomly selected admissions data are therefore: (1) restriction-of-range artifacts and (2) confounded variables.

### Measurement

Every measurement can be best regarded as an estimate which includes unknown components of two types of error: (1) unsystematic and (2) systematic. Unsystematic error randomly increases and decreases the measurement value which is observed (relative to the true value of the object or process being measured). Given a large number of measurements, however, the positive and negative errors tend to cancel, leaving a mean value that is virtually the true mean for that set of measurements. Systematic error affects the recorded measurement value in a consistent way (such as always mistakenly using a meterstick instead of a yardstick); regardless of the number of measurements taken, the error remains in the computed mean as well as in the individual measures. Nevertheless, if the nature of the systematic error comes to light, individual measurements and group statistics may be corrected. Measurement error generally refers to the random kind of error, whereas, systematic error in the measurement is an unaccounted for factor which has a nonrandom influence on the observed scores. If the systematic error is due to instrumentation or procedures, the factor may be called an "artifactual factor". Otherwise, the systematic factors

will be attributed to uncontrolled variables in the real data. Of course, errors also differ in level (or magnitude); error that doesn't differ in magnitude across samples is known as identically distributed error (unilevel), whereas error that does differ across samples (is multilevel) is known as moderated error.

In the context of a distribution, level of error is known as error variance. If error variance is multilevel (or heterogeneous; variance differs across factor levels), it is said to be moderated by that factor. Where error variance is multilevel and, in fact, correlated with the levels of the factor, the error is systematic-- a special case of moderated error.

Where it can be determined that error has systematic or random qualities, the possibility of controlling the error becomes more feasible.

### The Reliability of Validity Coefficients

The effect of error on correlation coefficients is more complex than is its effect on observed scores. With no error, the bivariate correlation is a consistent maximum: the coefficient of the two latent traits. To use a biological analogue, error may be likened to a parasite that invades a "host" variable. From a maximum, latent-trait correlation value, coefficients decline in value as greater levels of error affect the observed scores. This is always true for random error and is virtually always

true for systematic error (the exceptional cases being (1) where error adds a constant value to its host variable, or (2) where systematic error is perfectly correlated with its host variable).

Where the error in the observed scores is random, or where the error is systematic relative to its host variable, an unbiased estimate of the expected population coefficient can be computed. It can be computed with precision, moreover, if from a large sample; of course, the resulting correlation coefficient will be attenuated from the latent trait coefficient value due to the random error. Where error varies systematically relative to external influences, however, the computed estimate of the expected population coefficient may be inaccurate in some consistent fashion (biased). The biased estimate of the correlation of the latent traits would, therefore, require a correction of the observed-score correlation.

It is common for error to have attributes of both systematic and random error. It may appear to be normally distributed as in the case of random error, yet also prove to be reducible by the addition of variance controls. This would be the case of a moderating variable (such as year or location) where error may vary across units of the variable (e.g. times or sites) in either a systematic or a random fashion. By blocking on potential moderator variables (see Neter, Wasserman, & Kutner, 1985) or by using other

statistical controls, the moderated error can be removed or reduced.

The restriction-of-range problem is analogous to the problem of unreliability. Both can be accounted for in terms of the proportion of true variance to error variance. Both unreliability and range restriction are reflected in coefficients which are reduced when the proportion of error variance increases. By reducing the range of the variation in the sample of scores (as a consequence of selecting candidates via a cut-score criterion), the proportion of true variance is decreased, and, as a consequence, the proportion of error variance is directly increased (even with no change in the amount of error variance). If the error variance is substantially eliminated, the coefficient approaches the latent-trait value (in the case of parallel tests, that value should be one, though in the case of latent traits, the value could range between positive one and negative one). Linn and Hastings, (1984, p.166) provide a good discussion of the range restriction issue. Variation in range only affects raw-score regression coefficients when the dependent variable range is subject to variation between the calibration and the application samples (Richards, 1982), although the precision of this unbiased estimate depends heavily on a large sample size.

Richards (1982) discusses the data characteristics displayed in Table 3, which result in error artifacts under (1) raw-score and (2) standardized regression/

correlation. Most notably, raw-score regression coefficient estimates are unbiased by measurement error in the dependent variable, whereas standardized regression coefficients are unbiased by variation in scale units. Neither type of regression is immune from bias due to measurement error in the independent variable.

Although artifactual error in computed statistics may often be reduced through the use of various correction formulas (e.g. for unreliability or for range restriction), these corrections, nevertheless, are limited by the analyst's ability (1) to identify the affected variables (2) to determine levels of variance or reliability under other circumstances. Furthermore, a "corrected" coefficient cannot be assumed to be completely accurate, and may be expected to be conservative (see Linn, Harnisch, & Dunbar, 1981).

Table 3

Regression bias factors based on Richards (1982)

BIAS FACTOR	TYPE OF REGRESSION	
	Raw-score	Standardized
Units of measure differ	XX	
Dispersion of independent variable		XX
Unreliability of dependent variable		XX
Dispersion of dependent variable	XX	XX
Unreliability of independent variable	XX	XX
Norm referent scale	XX	XX
Change in test length	XX	XX
Selection on meditating variable	XX	XX

Correction of range restriction for a standardized regression equation requires the correction of each partial coefficient. Correction of the raw-score regression model is difficult because it requires the generation of a constant in addition to the transformation of partial coefficients into "b" weights (raw score coefficients).

Some important implications of this theory for selection need to be considered:

(1) If a sample is not randomly drawn its statistics will, nevertheless, represent its population as a whole if all of its relevant characteristics are invariant from member to member (for example, all Girl Scouts are invariant with respect to gender and relatively invariant with respect to age).

(2) If a sample is not randomly drawn, its statistics will also represent its population if the sample is large enough and if the subjects happen to be representative; automobile drivers are random relative to gender and political party preference; ten cars in a line that are picked from a public parking lot may not reflect population composition accurately, but a few hundred cars picked as a block from a parking lot may represent population composition quite accurately relative to gender or political preference. However, sample correlations are apt to be biased due to moderated error (non-random samples tend to systematically select certain subgroups).

(3) In practice, artifactual differences can be anticipated. Because many factors may be predictive of a particular kind of human performance, one must assume that (a) people may perform at a similar level even though they differ with respect to particular attributes (abilities on several factors may compensate for deficiencies on other factors); and (b) for a given year or location, non-random pressures (self-selection or other non-random selection) must be expected to favor particular factors/attributes resulting in samples which are systematically different from the population as a whole. For instance, a change in requirements for admission to human medicine programs may affect the rate and quality of applications to veterinary medicine.

Sample subgroups may differ in quality and level of preselection prior to inclusion in the sample, due to either self or institutional selection. Aggressive students may be over-represented due to self-selection, and range widely on required aptitudes while students with high verbal skills may be over-represented due to institutional selection and they may range very little on required aptitudes. The effects of these disparities are (a) to create the appearance of a differential validity of any fixed selection formula for the various applicant subgroupings (e.g. verbal, aggressive) and (b) to create the appearance of a differential selection formula validity across samples (e.g. applicants of different years or



locations may differ in their subgroup structure, see Linn, 1983; and Linn & Hastings, 1984). Nevertheless, the validity difference would be largely artifactual, a consequence of range- restriction due to selection.

Differential validity may actually exist independent of the artifactual manifestation, of course; however, a more insightful conceptualization is to attribute this particular validity problem to model misspecification (i.e., a selection formulation lacking in one or more important variables, such as origin of UGPA).

In short, under the uncontrolled conditions of an investigation-level study such as the analysis of graduate admissions data, the likelihood of moderated/ confounded prediction across years or locations is substantial. The result of moderated prediction may be biased regression and validity coefficients.

### Multiple Regression

The usual mathematical procedure used in multiple regression yields what is called the ordinary least squares (OLS) estimate of the criterion (or simply the least squares estimate). This term means that the sum of the squared prediction errors is minimized for the data used (see Neter, Wasserman, & Kutner, 1985). The OLS linear model that is calibrated, however, is truly OLS only with respect to the specific combination of the calibration predictors and the calibration criterion. The OLS

correlation (coefficient=  $R_{OLS}$ ) is the optimal correlation of predicted and actual scores because of the following: (1) scores are calculated by applying the calibrated linear model back onto the calibration predictors, (2) the formula (the linear model) was specifically developed to predict the same criterion data; the subsequent correlation of "ideal" criterion and "ideal" estimated criterion scores is optimal, (3) the OLS correlation is apt to be inflated, to some extent, due to the chance correlation of error with the criterion; when this chance correlation melts away during the application of the linear model to a new sample, the decrease in the coefficient upon cross- validity measurement is called shrinkage.

Multiple regression assumes that (1) responses on the dependent variable are independent, that (2) the variance is constant across cases, that (3) the errors are normally distributed, that (4) the system being modeled is in a steady-state equilibrium, and that (5) errors are uncorrelated (Kenny, 1979, pp.50, 51).

If all of the predictive factors are perfectly represented by the set of predictors and the criterion, the calibrated model will be optimal (although only a perfect correlation if in a totally determined situation where the latent trait correlation is one). Otherwise, as Deegan (1976) and Pedhazur (1982) explain, the following will be true: (1) where superfluous factors are included among the predictors (Deegan's overspecified model), unsystematic

error will be added to the predicted scores when the model is applied, thus causing an underestimate of the validity coefficient of the prediction scores (attenuation due to unreliability); (2) where some factors are omitted from the set of predictors (Deegan's underspecified model which includes the case of omitted independent factors; Deegan, p. 238), systematic error will be added to the predicted scores (this problem can be overcome only by providing the missing predictor data); (3) when a combination of these two situations exists, the model is said to be misspecified (misspecified models have biased parameter estimates which exhibit an interactive character, Deegan, p.238; obviously, without evidence that all predictive factors are appropriately represented, all practical models must be assumed to be somewhat misspecified); (4) where important predictor variables are highly correlated, multicollinearity is said to exist.

Pedhazur (1982) points out that there is differential use of the term, multicollinearity, but its problematic manifestation is biased predictor weights. This problem results in systematic error being added to predicted scores when the calibrated model is applied to non-calibration data (new data). Much work has been done to develop alternate regression techniques to cope with the problem of multicollinearity. Unfortunately, most of these techniques are helpful only in the most severe circumstances (Huberty & Mourad, 1980; Morris, 1986; Cattin, 1981). Kenny (1979)

lists three characteristics associated with low predictor error of measurement. These characteristics are also associated with reduced problems of multicollinearity: (1) a high reliability, (2) a low regression coefficient, and (3) a low predictor intercorrelation. He also adds that multicollinearity decreases as the number of predictors decreases relative to the number of cases.

If, therefore, certain data characteristics exist, the regression procedure will yield an equation which will specify an efficient means of weighting several variables in order to estimate a criterion. Regression equations may suffer from either too few or too many predictor variables. A lack of predictors results in biased regression coefficient estimates, while too many predictors makes the coefficient estimates less reliable. High predictor variable intercorrelation may also bias coefficient estimates although this problem is less severe for predictors with high reliability and/or a moderate to low correlation with the criterion.

#### Multivariate and Univariate Analysis of Variance and the T-Test

The t-test is a special case of the more general ANOVA; therefore they share similar theoretical assumptions. The ANOVA procedure yields a ratio of the variation of means to the variation of simple scores. Under the null hypothesis of no effects for the levels of a factor studied, the ratio

will be approximately one (1:1). Otherwise the variation of means will result in a ratio greater than one thereby suggesting the implausibility of the null hypothesis. The analysis of variance requires an interval-level dependent variable and a nominal-level independent variable with at least two levels (a one-way ANOVA). Where there is more than one independent variable, the ANOVA may be two-way, three-way, or etc. If two or more multi-level independent variables are in the ANOVA design it is classified as a factorial design.

The ANOVA procedure makes certain statistical assumptions of the data being analyzed; where these assumptions are violated, ANOVA findings may be less valid. Under all circumstances the dependent variable scores must be independent of each other. And where the sample-sizes differ per condition (cell-sizes), the variances must be equivalent. Violations of other assumptions tend to be less important (see Kirk, 1982, pp. 74-79).

Another way of looking at the issue of the independence of the dependent variable scores is in terms of accounting for variance. When sampling is less than random, dependent variable scores may not be independent. If dependent variable scores are correlated, there is apt to be a variable whose control would result in independent scores. The question becomes, therefore, whether or not the important factors have been controlled in the study's design. The answer to this question requires a rational

analysis of the potential causes of score variance and the adequacy of the study's design to differentiate such variation. For example, where the dependent variable is a performance score, it is critical that there be no overlap in respondents' between-condition samples. However, where the dependent variable is error-due-to-method, as in the present study, variance in the dependent variable will not be substantially affected by randomly overlapping samples of respondents whose scores are fixed prior to the experiment; the systematic error will be virtually a consequence of the mathematical transformations attributable to the methods. If, however, the sampling is restricted from certain levels of population variation (e.g. particular years, locations, ability levels), the restricted variables need to be included in the experimental design as independent variables; otherwise the dependent variable is likely to be dependent on one or more unspecified moderators and ANOVA validity will suffer.

Where cell variances are unequal, it is important that cell sizes be equal. If cell variances and sizes are approximately equivalent, the ANOVA validity should be acceptable, particularly where the sample sizes are large.

Extending single dependent variable analysis of variance to the multivariate case (MANOVA), additional assumptions must be met in order to make valid statistical inferences. Tabachnick and Fidell (1983, pp.231-235) include the following assumptions and requirements: (1)

homogeneity of covariances replaces its ANOVA analogue, equality of cell variances; (2) the number of cases per cell must exceed the number of dependent variables; (3) the dependent variables should exhibit a multivariate normal distribution; (4) there should be no outlier cases; (5) all dependent variables and covariates should share linear relations; and (6) dependent variables should exhibit an absence of multicollinearity.

In short, the t-test, ANOVA, and MANOVA, test between group variation by the variation of simple scores in order to conclude whether variation between groups exceeds limits acceptable for the null hypothesis. Independence of responses and the equality of either cell sizes or variance are the critical assumptions. As this procedure is extended to the multivariate situation, some additional requirements become important.

## CHAPTER III

### PROPOSED THEORY

The term  $Y_{\text{GGPA}_i}$  in Equation 1 represents the estimated GGPA (graduate school GPA) for student  $i$ . For a conventional GC (general-criterion) multiple regression calibration of pooled cross-year and/or cross-location samples, the selection formula is identical to equation (1) where GGPA has been regressed on mutual predictors (cross-year/cross-site calibration can utilize only those predictors which are mutually available: exceptional predictors must be discarded):

$$Y_{\text{GGPA}_i} = \beta_0 + \beta_{\text{UGPA1}} * X_{\text{UGPA1}_i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni} \quad (1)$$

Where,

$Y_{\text{GGPA}_i}$  = Estimate of Grad Program GGPA  
 $X_{\text{UGPA1}_i}$  = Undergraduate UGPA  
( $X_{2i} \dots X_{ni}$ ) = Other application variables

In practice, a formula is often calibrated with the GGPA of only the first year. For the GC (general-criterion) approach, the predictor data of non-admitted applicants is ignored, while only the predictors of accepted students are saved. The predictors lie idle through the freshman through junior years, until the end of the senior year when the final GGPA is available as a criterion (C). The formula is calibrated on a selected



range of applicants in year four, nevertheless, applied to the full range of applicants in year five.

In the LP (local-proxy) regression approach, PVUGPA, a subset of UGPA (specifically, the college pre-veterinary UGPA) becomes the proxy-criterion for the solitary-year-and-location sample. Assuming that  $PVUGPA = GGPA_1 + \text{error}$ , equation (1) also applies to the LP calibration when the UGPA subset serves as a proxy criterion and when raw-score regression is used with an adequate sample size (error in the dependent variable does not bias the raw-score regression coefficient; Richards, 1982). In contrast to the conventional approach, however, the LP calibration cases are also the cases to which the subsequent selection formula is applied. A regression model using all applicant data for year "Y" is calibrated at the time of application, using a subset of UGPA as the criterion (PVUGPA, the UGPA for the veterinary prerequisites) and UGPA as one of several predictors. The resulting regression equation is used as the selection formula for the same set of year "Y" applicants; the selection formula is applied to the year "Y" applicant data to compute selection scores for each applicant.

#### General-criterion (GC) Regression Approach

The conventional strategy for implementing multiple regression in the development of a selection formula will be identified as the general-criterion approach, or GC.

For this model, a large pool of cases are cumulated across years and/or across locations adding error where years and/or locations are moderators. An authentic though restricted criterion is used (e.g. GGPA; it represents mostly the higher performing applicants). The optimal formula for predicting the criterion is limited by (1) the availability of mutual predictors among all the cases, and by (2) the predictiveness of those variables for that particular pool of cases. Once a selection formula is calibrated, the permissible predictor variables and their accompanying weights are set until the next calibration. Depending on the range of years and locations represented in the calibration sample, the formula may be generalizable across time and locations. If the selection-rating system changes over time or location, however, the potential validity of the formula may decline. A large number of assumptions are required to support this approach due to potential moderator variables, multicollinearity, and restricted range problems. Moderators and multicollinearity become important concerns because the formula is being generalized to cases outside the calibration pool (usually, across years and/or locations). Because calibration case UGPAs are range-restricted relative to the applicant pool, corrections for restriction-in-range are required to adjust the calibrated selection formula.

With this approach, error may enter by way of the following channels:

- Ca. criterion variables
- Cb. predictor variables
- Cc. statistical artifact (connected with the multiple regression procedure or correction specifications)
- Cd. individual effects
- Ce. moderators (e.g. year or location effects)

Potential sources of systematic error include the following:

- Sa. individual aptitude/motivation variation
- Sb. halo and other individual error
- Sc. qualitative/quantitative metric variation
- Sd. restricted content domain
- Se. unspecified predictors
- Sf. multicollinearity among predictors
- Sg. individual effects
- Sh. moderator effects (e.g. years and locations)

The notion of an individual effect being systematic may seem dubious to some. Nevertheless, it is both possible for individual error (such as a halo effect) to (1) occur across graders in a consistent fashion or, (2) to affect graders in a random fashion. Potential sources of random error include the following:

- Ra. measurement error
- Rb. sampling error
- Rc. individual effects
- Rd. moderator effects (e.g. years and locations)
- Re. superfluous predictors

Some advantages for the general approach include the following:

- Aa. criterion accuracy--use of the authentic criterion
- Ab. potential generalizability across years/or locations  
(diminishing the need for frequent recalibrations)
- Ac. the accumulation of a large calibration pool will  
diminish the problems of sampling error relative to  
the estimation of a population observed-correlation,  
given that (1) sampling is equivalent to random  
across years and locations and (2) moderator effects  
are largely absent

Disadvantages of the general approach include the following:

- Da. the need for previous cohort data
- Db. a fairly complicated analysis procedure is required
- Dc. the selection formula is fixed (closed to new  
predictors)
- Dd. the criterion exists only in a selected sample, thus  
requiring corrections of the calibrated selection  
formula due to restrictions-of-range

- De. the dangers of systematic error due to individual effects, year effects, and/or location effects (compounded since years and locations are seldom drawn randomly or even with large numbers)
- Df. the predictor pool is diminished because some locations or years don't have conforming variables, thus increasing the underspecification of predictors problem (and systematic error)

Optimum conditions for the use of the general approach are as follows:

- Oa. a large calibration sample
- Ob. a large application sample
- Oc. the stability of qualitative/quantitative metrics of selection variables across locations
- Od. multiple independent (orthogonal) variables which closely predict the criterion
- Oe. a rich pool of parallel predictors which exist across locations
- Of. a minimum drift of the population model over time
- Og. a high validity of population model over locations
- Oh. stable demographic characteristics
- Oi. a stable applicant pool (despite recruitment variation)

### Using Multiple Regression to Shrink Error

If a predictor such as UGPA has been measured with a variety of attribution rules (for rating performance) across applicants, the pooled predictor values will include error moderated by locations and/or years. Linn (1966) demonstrated how raw-score multiple regression can be used to "shrink" (reduce in magnitude) the moderated error. This can be done where the following appropriate conditions exist: (1) there are many cases sharing a given rule, (2) at least two mutual measures of performance are known to be standard for all of the applicants, and (3) these mutual measures of performance are similar in nature to the uncorrected predictor (measures share common factors). For example, to correct (partially) UGPAs, one would like to have a set of cohort data where, in addition to (1) the uncorrected UGPA, there is (2) a mutual GGPA (graduate program GPA), (3) a mutual admissions test score and (4) a dummy variable for each rule (or school of origin). By using the graduate program GGPA as the calibration criterion and the test score and uncorrected college UGPA as predictors, an adjustment weight can be obtained for correcting cases of a similar rule.

If college UGPA, test score, and graduate program GGPA are parallel measures, then the predicted scores resulting from the application of the calibrated raw-score multiple regression model, should be estimates of graduate GGPA. This constitutes a particular case of improving an

underspecified model, since school variables were correlated with the criterion and accounted for a certain type of variation which otherwise would have been regarded as error. The following special cases are possible modifications of Linn's model augmentation approach:

A. The case where the school variable is absent:

If, however, the raw score multiple regression is done without variables identifying the school of origin, the predicted scores will be estimates of graduate GGPA plus some moderated error due to the absence of the missing predictors (school variables). Nevertheless, to the extent that the variation in rule (of UGPA standards) is random and the number of cases is sufficiently large, the predicted scores may be partly-corrected (errors of estimate would be unbiased).

B. The case where the sample has a predominate rule:

If a substantial proportion of the cases already have a similar rule, the correction problem, obviously, is diminished, and the precision of prediction for that particular rule-subgroup will increase. In contrast, subjects with non-conforming rules will be predicted with less accuracy. To the extent that the non-conforming rules differ randomly from the typical rule, quantitatively, and to the extent that qualitative variation is random in nature relative to the typical rule, the calibrated selection formula will be optimal for the whole of the

assorted rule-subgroups despite its inferior prediction for individuals having non-conforming rules. Where a predictor variable's rules vary greatly (lacking a predominate rule-subgroup), general precision will suffer and the calibrated selection formula will tend to select more for general as opposed to specific ability. This is because the predictor variable will be less reflective of specific ability due to varying qualitative and quantitative standards/ rules; hence, only general ability will tend to remain intact as a common factor.

C. The case where rules vary within the school:

It may be assumed that cross-institutional variation in academic standards is an important factor in criterion integrity. However, variation at other levels may be of equal or greater importance, such as at the curriculum or major level (Elliot & Strenta, 1988; McCornack & McLeod, 1988). On the other hand, academic-major variation may be due to factors independent of subject matter, such as specific course content or specific class instructor. Error at this level cannot be reduced by merely controlling for location.

D. Where prediction is improved for extreme cases:

Should a predictor differentiate unequally along some important dimension of a sample of cases, the resulting improvement in prediction would affect only a restricted range of cases (i.e. interview ratings may only be valid when augmenting prediction for the highest ability



students). In this instance an improvement in predictiveness does not improve prediction in the cut score region. Therefore, an increase in a coefficient value may not correspond to any real improvement in selection formula validity.

E. The case where a proxy criterion is used:

If the necessary conditions are obtained for the shrinking of variable error (as outlined above) except that the (graduate) GGPA criterion is replaced with a proxy (UGPA), the consequent model will estimate a GGPA with prediction error shrunken (relative to the accuracy of the proxy variable and the randomness of the school-rules in which the proxy variable is measured). The greater the potential year or location influence on the criterion, the greater the potential for reducing prediction error. In addition to allowing control of year or location influence, a proxy criterion may also be useful as a means of extending the size and variability of the calibration sample (the sample used to calibrate a selection formula).

If the criterion (GGPA) represents the same measurement factors as a proxy (UGPA) then the proxy may be regarded as  $c+e$  (criterion + error). If a predictor variable then correlates with the criterion, then  $r_{cp} > r_{(c+e)p}$ , since attenuation of correlation results from unreliability (or error) in a measure. However, if raw-score regression coefficients are used, then the criterion with error will combine with a predictor to yield an

unbiased estimate of the raw-score regression coefficient (Richards, 1982). Hence, to the extent that UGPA may approximate GGPA plus a random error component, and to the extent that prediction factors vacillate annually, or locally, a proxy criterion may improve prediction and selection.

#### Proxy-criterion Alternatives to Conventional Prediction

The use of a proxy criterion allows additional alternatives to the conventional procedure for calibrating a selection formula. The availability of a suitable proxy criterion may, potentially, extend the number of cases available to analysis in addition to extending the variability of the cases available. Depending on the quality of the proxy criterion, the sample may be controlled for year or location variability by restricting the sample on such confounding/moderating variables. If the proxy criterion allows an increase in the usable sample size per year, the loss of other-year or other-location cases may not be critical. Where number of cases seems to be a more critical factor than moderator problems, the sample may be increased by adding criterion-absent cases to those having a criterion, because a suitable proxy can substitute for the criterion. Should no ideal criteria be available, use of the above procedure with all cases while using an inferior criterion may still be beneficial. Tables 4 through 9 (on pages 62-65) contrast four

calibration methods on the following features: (5) potential channels for error, (6) potential sources of systematic error, (7) potential sources of random error, (8) potential advantages, (9) potential disadvantages, and (10) optimal conditions for use of each of four proxy methods. A discussion of the attributes of three potential calibration approaches precedes presentation of the tables (items are ordered according to the tables):

#### 1. Local-proxy (LP)

The local-proxy calibration uses a single-year, single-site sample to estimate predictor equation parameters. If a proxy variable is appropriate and available for replacing an unavailable criterion (e.g. future GGPA), a selection formula can be calibrated and applied back to the same data to yield prediction scores having a validity approaching optimal validity. Using raw-score regression and assuming that the proxy approximates  $c+e$  (criterion + random error), the linear model generated will approximate the OLS regression model. Validity here depends greatly on the quality of the proxy criterion, although a second important asset would be a rich assortment of reliable predictors. Error, therefore, may enter by way of the following channels:

Ca. criterion variables (particularly, the proxy variable)

Cb. predictor variables (present or absent)

Cc. a statistical artifact (connected with the multiple regression procedure)

Ce. moderators (e.g. gender, social class)

Potential sources of systematic error would be the following:

Sa. individual aptitude variation

Sb. halo or other individual error

Sc. qualitative/quantitative rule variation

Sd. a restricted content domain

Se. unspecified predictors

Sf. the level of multicollinearity among predictors

Sg. individual motivation variation

Sh. moderators (e.g. gender, social class)

There is little reason to anticipate substantial changes in individual aptitudes over the course of a graduate program, although such is possible (e.g. brain disease or injury). More likely is the possibility of halo effects which consist in systematic increases or decreases in the criterion score due to subjective bias on the part of the instructor who assigns the predictor or criterion score. Although differences in grade attribution-rules among schools offer the potential for systematic error in predicting criterion scores using school UGPA as a predictor, Linn (1966) observed that this problem was insignificant when admissions-test data was

included among the predictors. Where some cases contain predictor scores representing performance on a narrower content domain, those cases are likely to be overpredicted on their criterion performance. Of course, a change in level of motivation may affect criterion performance. Where important predictors are excluded from the prediction model, systematic bias is added to the estimated criterion. Although multicollinearity is most problematic where a regression model is being generalized to additional samples, it nevertheless can play a minor role in LP estimation. In particular, multicollinearity may distort variable weights so that when the calibrated selection weight is applied to a parallel variable (i.e. a weight calibrated on mostly 1983 MCAT-- Medical College Admissions Test-- scores gets applied to a 1987 MCAT score), systematic bias may be added. This problem is exacerbated by the level of measurement error present. The absence of important predictor variables from the selection model will also distort the calibrated selection formula. Lastly, the accumulation of moderators is likely to be a consequence of non-random sampling. Moderators, in turn, may have a systematic influence on prediction error.

Sources of unsystematic error in the predictors would likely bias the selection formula. Random error affecting the proxy criterion may distort regression weights for low sample-size-to-measurement-error ratios while unreliable predictors violate the regression assumption of perfectly

measured independent variables. Sources of random error would be the following:

- Ra. measurement error
- Rb. sampling error
- Rd. moderators (e.g. gender, social class)

Like measurement error, sampling error is defined by statisticians as random error, although statistical differences between random samples may be partly systematic (Pedhazur, 1982, is an exception who includes systematic error as a type of measurement error). The fact that statisticians prefer to attribute the systematic component of sampling error to unspecified predictors rather than to the pool of sampling error does not alter the practical fact that differences between sample statistics will always be partly due to the problem of unspecified predictors. Moderator error can be expected to be random (i.e. level of measurement error in MCAT scores may vacillate randomly across years), except where a theoretical basis exists for a systematic nature.

Some advantages for the LP approach include the following:

- Ad. a simple analysis procedure
- Ae. freedom from the need for previous cohort data
- Af. the proxy criterion provides the desirable  
feature of interval level scale where some graduate  
programs have only dichotomous grading (pass/fail)

- Ag. the option of adding predictor variables for any new set of applicants (e.g. alternate admission test scores can be specified and weighted)
- Ah. freedom from some potential systematic or random ability change) year effects, and location effects
- Ai. range restriction problems are largely eliminated by the implementation of all applicant cases with proxy criteria
- Aj. non-admitted cases can be used

Disadvantages of the LP approach include the following:

- Dg. the need to recalibrate a new selection formula for each set of new applicants
- Dh. the danger of individual aptitude or motivation change during the course of the program in question
- Di. potential proxy criterion inadequacies

Although any change in individual aptitude or motivation would decrease the validity of an aptitude measurement (lower the regression coefficient corresponding to the aptitude measure), where an authentic criterion is used, the lower validity would be accurate. With use of a proxy criterion, however, the validity would remain inflated because the proxy would not reflect the problematic trait variation.

Conditions under which the LP approach will be optimum:

- Oa/b. a large calibration/ application sample
- Oc. the stability of qualitative/quantitative metrics  
of selection variables across locations
- Od. multiple independent (orthogonal) variables which  
closely predict the criterion
- Oe. a rich pool of parallel predictors which exist across  
locations
- Oj. the stability of aptitudes and motivation

## 2. General-proxy (GP)

The general-proxy (GP) approach uses a multi-year sample and a proxy criterion. With the pooled sample fixed relative to location, the proxy criterion approach can be used to calibrate a selection formula from a pool of cases accumulated across several years. The advantage of this approach is the potential for compensating for cases lost while controlling for a moderator, because it allows a greater number of usable cases within each applicant-year sample. It may also generalize across years, thus reducing the frequency of the need to recalibrate a formula. Relative to the local approach, a potential liability is the possible systematic error due to year effects. Because only a small range of years of data is likely to be accessible, the external validity of the selection formula may be poor (the local approach does not attempt to generalize). Other variations of this compromise approach



are possible also, such as using a pooled sample fixed relative to year but not to location. Error may enter by way of the following channels:

- Ca. criterion variables (particularly, the proxy variable)
- Cb. predictor variables
- Cc. statistical artifact (connected with the multiple regression procedure)
- Ce. moderators (e.g. year or location)

Potential sources of systematic error include the following:

- Sa. individual aptitude variation
- Sb. halo and other individual error
- Sc. qualitative/quantitative rule variation
- Sd. a restricted content domain
- Se. unspecified predictors
- Sf. the level of multicollinearity among predictors
- Sg. individual effects
- Sh. moderator effects (e.g. year or location effects)

Potential sources of random error would include the following:

- Ra. measurement error
- Rb. sampling error
- Rc. moderator effects (e.g. year or location effects)

**Rd. individual effects**

**Re. superfluous predictors**

Some advantages for the general-proxy approach include the following:

**Ab. generalizability to other locations and/or years**

**Ac. sample size can be increased by pooling**

**Ad. a simple analysis procedure**

**Af. the proxy criterion provides the desirable**

**feature of interval level scale where some graduate programs have only dichotomous grading (pass/fail)**

**Ah. freedom from potential moderated error due to**

**location effects (or alternatively, freedom from moderated error due to year effects)**

**Ai. range restriction problems are largely eliminated**

**by the implementation of all applicant cases with proxy criteria**

**Aj. allows use of non-admitted cases**

Disadvantages of the general-proxy approach include the following:

**Da. the need for previous cohort data**

**Dc. the selection formula is fixed**

**De. error due to years, sites, and individuals**

**Df. restrictive tendency in predictor pool**

- Dh. the danger of individual aptitude or motivation change during the course of the program in question
- Di. proxy criterion inadequacies

Conditions under which the general-proxy approach will be optimum:

- Oa/b. a large calibration/ application sample
- Oc. the stability of qualitative/quantitative rules of selection variables across locations
- Od. multiple independent (orthogonal) variables which closely predict the criterion
- Oe. a rich pool of parallel predictors which exist across locations
- Of. the stability of population model across years
- Oj. the stability of aptitudes and motivation

### 3. General-criterion-proxy or General-mixed (GM)

The general-mixed (GM) calibration approach includes a multi-year sample and a conditional criterion: an authentic or a proxy criterion. This is the conventional strategy for implementing multiple regression in the development of a selection formula except with the inclusion of non-admitted graduate applicant cases. Non-admitted graduate cases utilize PVUGPA (prerequisite veterinary course UGPA) as a proxy for the authentic graduate GGPA criterion. A large pool of cases are cumulated across years and/or across locations. The optimal formula for

predicting the criterion is limited by the predictiveness of those variables for that particular pool of cases. Once a selection formula is calibrated, the permissible predictor variables and accompanying weights are set until the next calibration. Depending on the range of years and locations represented in the calibration sample, the formula may be generalizable across time and locations. If the system changes over time or location, the potential validity of the formula may decline. As in the conventional GC (general-criterion) model, a greater number of assumptions are required to support this approach, but the range restriction problems are effectively resolved so that corrections for selection may be unnecessary. Multicollinearity remains a concern since the formula is being generalized to cases outside the calibration pool and, usually, across years and/or locations.

With this approach, error may enter by way of the following channels:

- Ca. criterion variables
- Cb. predictor variables
- Cc. statistical artifact (connected with the multiple regression procedure or correction specifications)
- Cd. individual effects
- Ce. moderators (e.g. year and location)

Potential sources of systematic error include the following:

- Sa. individual aptitude/motivation variation
- Sb. halo and other individual error
- Sc. qualitative/quantitative metric variation
- Sd. a restricted content domain
- Se. unspecified predictors
- Sf. multicollinearity among predictors
- Sg. individual motivation variation
- Sh. moderator effects (e.g. year or location effects)

Important sources of potential random error include the following:

- Ra. measurement error
- Rb. sampling error
- Rc. individual effects
- Rd. moderator effects (e.g. year or location effects)
- Re. superfluous predictors

Some advantages for the GM approach include the following:

- Aa. criterion accuracy- use of the authentic criterion
- Ab. potential generalizability across years/or  
locations (diminishing the need for frequent  
calibrations)
- Ac. the accumulation of a large calibration pool will  
largely diminish the bias effect of sampling error  
given that sampling is equivalent to random across  
years (or locations)

**Ad. simplicity of analysis**

**Ai. range restriction problems are largely eliminated  
by the implementation of all applicant cases with  
proxy criterions**

**Aj. increased sample size due to added rejectee cases**

**Disadvantages of the general-criterion-proxy approach  
include the following:**

**Da. the need for previous cohort data**

**Dc. the formula is fixed (is closed to new  
predictors)**

**De. dangers of systematic error due to individual  
effects, year effects, and/or location effects  
(compounded since years and locations are seldom  
drawn randomly or even in large numbers)**

**Df. the predictor pool is diminished as some  
locations don't have conforming variables thus  
increasing the underspecification-of-predictors  
problem and its systematic error**

**Dh. the danger of individual aptitude or motivation  
change during the course of the program in question**

**Di. proxy criterion inadequacies**

**Optimum conditions for the use of the GM approach:**

**Oa. a large calibration sample**

**Ob. a large application sample**

- Oc. the stability of qualitative/quantitative rules  
of selection variables across locations**
- Od. multiple independent (orthogonal) variables which  
closely predict the criterion**
- Oe. a rich pool of parallel predictors which exist across  
locations**
- Of. a minimum drift of population model over time**
- Og. a high validity of population model over locations**
- Oh. stable demographic characteristics**
- Oi. a stable applicant pool (despite recruitment  
variation)**
- Oj. stability of aptitudes and motivation**

It is expected (in any measurement situation) that unspecified factors will add a random distribution of error to the scores of the cases being measured. Where scores are estimates of future ratings (predicted scores) it is possible to actually obtain a measure of these errors in order to study the nature of the error. This is done by subtracting subsequent outcome scores (the criterion) from the prediction scores. If these errors are partly correlated with one or more potential predictors, they are systematic and it is possible that the score prediction formula may be improved by modifying predictors or their weights. If the errors appear to be non-randomly distributed for a large sample, but they fail to correlate with conceivable predictors, there nevertheless is likely

to be an unspecified factor of predictive importance; therefore the selection /prediction formula will be inaccurate. If the distribution of errors is random, there may, nevertheless, be a moderating variable (e.g. years, locations) for which the level of error changes in an unsystematic way. This also results in a selection/prediction formula which is inaccurate. Only where error is randomly distributed and apparently irreducible by (1) the addition of predictors or by (2) controlling for potential moderator variables, can it be concluded that the selection/prediction formula is precise.

Table 4

Channels by which error may affect calibration methods

CHANNELS FOR ERROR	GC	LP	GP	GM
Ca Criterion variables . . . . .	XX	XX	XX	XX
Cb Predictor variables . . . . .	XX	XX	XX	XX
Cc Statistical artifacts . . . . .	XX	XX	XX	XX
Cd Individual effects . . . . .	XX			XX
Ce Moderators . . . . .	XX	XX	XX	XX

GC= General-Criterion prediction method

LP= Local-Proxy prediction method

GP= General-Proxy prediction method

GM= General-Mixed prediction method



Table 5

## Potential systematic error for each calibration method

SOURCES OF SYSTEMATIC ERROR	GC	LP	GP	GM
Sa Aptitude/ motivation variation . . . . .	XX	XX	XX	XX
Sb Halo and other individual error . . . . .	XX	XX	XX	XX
Sc Scale irregularity . . . . .	XX	XX	XX	XX
Sd Restricted content domain . . . . .	XX	XX	XX	XX
Se Unspecified predictors . . . . .	XX	XX	XX	XX
Sf Multicollinearity among predictors . . . . .	XX	XX	XX	XX
Sg Individual effects . . . . .	XX	XX	XX	XX
Sh Moderator effects . . . . .	XX	XX	XX	XX
GC= General-Criterion prediction method				
LP= Local-Proxy prediction method				
GP= General-Proxy prediction method				
GM= General-Mixed prediction method				

Table 6

## Potential random error for each of four calibration methods

SOURCES OF RANDOM ERROR	GC	LP	GP	GM
Ra Measurement error . . . . .	XX	XX	XX	XX
Rb Sampling error . . . . .	XX	XX	XX	XX
Rc Individual effects . . . . .	XX		XX	XX
Rd Moderator effects . . . . .	XX	XX	XX	XX
Re Superfluous predictors . . . . .	XX		XX	XX
GC= General-Criterion prediction method				
LP= Local-Proxy prediction method				
GP= General-Proxy prediction method				
GM= General-Mixed prediction method				

Table 7

## Potential advantages for each of four calibration methods

ADVANTAGES	GC	LP	GP	GM
Aa Criterion accuracy . . . . .	XX			XX
Ab Generalizability . . . . .	XX		XX	XX
Ac Increase sample by pooling . . . . .	XX		XX	XX
Ad Simplicity of analysis . . . . .		XX	XX	XX
Ae Needs only one year's admission data . .		XX		
Af Interval scale despite pass/fail GGPA .		XX	XX	
Ag Admits new predictors . . . . .		XX		
Ah Reduced year, site, and individual error .		XX	XX	
Ai Avoids range restriction . . . . .		XX	XX	XX
Aj Uses non-admitted cases . . . . .		XX	XX	XX

GC= General-Criterion prediction method

LP= Local-Proxy prediction method

GP= General-Proxy prediction method

GM= General-Mixed prediction method

Table 8

## Disadvantages of each of four calibration methods

DISADVANTAGES	GC	LP	GP	GM
Da Need for previous cohort data . . . . .	XX		XX	XX
Db Complicated analysis . . . . .	XX			
Dc Selection formula is fixed . . . . .	XX		XX	XX
Dd Selection on the criterion . . . . .	XX			
De Error due to years, sites, individuals .	XX		XX	XX
Df Restrictive tendency in predictor pool .	XX		XX	XX
Dg Need to recalibrate each year . . . . .		XX		
Dh Risks aptitude or motivation change . .		XX	XX	XX
Di Potential proxy criterion inadequacies .		XX	XX	XX

GC= General-Criterion prediction method

LP= Local-Proxy prediction method

GP= General-Proxy prediction method

GM= General-Mixed prediction method

Table 9

Optimal conditions for use of each calibration method

OPTIMAL CONDITIONS FOR METHOD	GC	LP	GP	GM
Oa Large calibration sample . . . . .	XX	XX	XX	XX
Ob Large application sample . . . . .	XX	XX	XX	XX
Oc Stability of cross-site scales . . . . .	XX	XX	XX	XX
Od Multiple, sound predictors . . . . .	XX	XX	XX	XX
Oe Cross-site predictors . . . . .	XX	XX	XX	XX
Of Temporal stability of model . . . . .	XX		XX	XX
Og Model validity across samples . . . . .	XX			XX
Oh Stable demographic characteristics . . . . .	XX			XX
Oi Stable applicant pool . . . . .	XX			XX
Oj Stability of aptitudes and motivation . . . . .		XX	XX	XX
GC= General-Criterion prediction method				
LP= Local-Proxy prediction method				
GP= General-Proxy prediction method				
GM= General-Mixed prediction method				

## Summary

The conventional regression approach (GC) to calibration of selection formulas may incorporate several usages of a proxy criterion in the estimation process, resulting in alternative regression procedures which are labeled LP, GP, and GM (LP= formula calibration from a single site and time, GP= calibration across sites or times using a proxy criterion, and GM= calibration across sites or times mixing both authentic and proxy criterion use). Tables 4 through 9 attempt to compare and contrast the methods with respect to their corruptibility, strengths, and conditions for optimal performance.

## CHAPTER IV

### DESIGN

#### Population Sample

The non-random sample consisted of five cohorts; four graduation cohorts and an additional three-year cohort from the College of Veterinary Medicine at Michigan State University. The cohorts included both accepted and rejected applicants. Applicants who obtained a vet school GGPA were considered to be GRADS (this includes the three-year cohort); all others were considered to be non-graduates, or NGRADS. Student attrition was estimated to be at less than two students per cohort. Some overlap existed among the cohorts, as some cases were repeat-applicants. While the annual number of applicants changed notably over the years, the number of candidates in-program was relatively constant (see Table 10). All applicants in this sample received scores on the New Medical College Admissions Test (MCAT) and were ranked for admissibility by a selection formula decided upon by the Veterinary College's admission committee. The high admissibility of an applicant could often mean that such an applicant might accept another candidacy elsewhere, therefore, some NGRADS were of this caliber.

Table 10

The study sample: Five cohorts of veterinary applicants

APPLICATION YEAR	GRADUATION YEAR	NUMBER OF APPLICANTS	NUMBER OF GRADUATES
1981	1985	327	90
1982	1986	327	89
1983	1987	310	103
1984	1988	273	88
1985	1989	248	101

Three graduation cohorts (1985, 1986, 1987) were used in the calibration sample for the GC, GP, and GM conditions and the other two cohorts (1988, 1989) were used for validity testing of GC, GP, GM, and LP conditions. For the LP condition, the same cohorts (1988, 1989) were used for both calibration and validity testing. The validity test weighted higher a portion of the graduating subset of the two validity-test cohorts; specifically, the validity-test used cases found in the lower third of the GRAD program GPA range. These cases were selected for their likely proximity to the admissions cut-score region. The higher performing students would be less affected by a different admissions cutting criterion.

### Predictors

Predictor variables were of two types, (1) within-unit predictors (which were the following ordinary predictor

variables: CUMGPA (cumulative UGPA), PVUGPA (veterinary prerequisites UGPA), honor points, prerequisite course honor points, credits without pass/fail credits, average credits per term, summer credits, total credits, number of terms, pass/fail credits, the Medical College Admissions test subtest scores, age, sex, interview scores I and II, work experience, veterinary experience, narrative work sample, and source of UGPA), and (2) between-unit variation (year). Although most of the variables were taken directly from admissions documents, the source-of-UGPA variable is defined especially for this research. This variable is defined from a retrospective analysis of student numbers together with consultation with an admissions staff-member to determine which applicants clearly were not MSU undergraduates. Subsequently, a small proportion of the applicants actually fell into the non-MSU category; this is particularly true for applicants who graduated from the program. Where applicant cases had missing values, default scores were assigned either by entering an average score or by entering a minimum score. This was similar to the practice of admission departments in determining an applicant's merit (see default score list in Appendix I).

Predictors were available from admissions data variables. Because many of these variables were not of interval scale, and because some of them were composites of several variables, such variables were deleted. Composites were sacrificed for single variables. The twenty-two

Table 11

Predictors used in levels of the intercorrelation factor

PREDICTORS	INTERCORRELATION LEVELS	
	>.45	<.45
cumulative UGPA . . . . .	XX	XX
veterinary prerequisites UGPA . .	XX	XX
veterinary prerequisites points .	XX	
honor points . . . . .	XX	XX
credits . . . . .	XX	
average term credits . . . . .	XX	XX
sum of pre-veterinary credits . .	XX	XX
total credits . . . . .	XX	
number of terms . . . . .	XX	
pass/fail credits . . . . .	XX	XX
MCAT Biology . . . . .	XX	
MCAT Chemistry . . . . .	XX	XX
MCAT Physics . . . . .	XX	
MCAT Quantitative. . . . .	XX	
MCAT Reading . . . . .	XX	XX
MCAT Science . . . . .	XX	
age . . . . .	XX	
sex . . . . .	XX	XX
veterinary experience . . . . .	XX	
work experience . . . . .	XX	XX
activities and achievements rating	XX	XX
narrative writing sample rating .	XX	XX

variables included in the predictor intercorrelation levels are listed in Table 11.

### Criteria

Program GGPA served as the authentic criterion. The proxy criterion (used with the LP, GP, and GM approaches) was the PVUGPA (veterinary prerequisite course UGPA).

## Analyses

All research hypotheses were addressed by analyses of either of the following two analyses: one primarily used to test calibration methods and another primarily used to test for calibration factor and year effects:

### Test of Methods (MANOVA<sub>TOM</sub>)

Sample. Three graduation cohorts were used for the calibration of GC, GP, and GM formulas with two additional cohorts reserved for the purpose of validity testing. The calibration of the LP formula used one of the reserved cohorts in each of its two calibrations. Validity testing for all calibrated formulas was on the two reserved cohorts. Two years' data (1981 and 1983) were drawn for the small sample GC, GP, and GM calibrations while three years' data were used for their large sample calibrations; the same two years were used as the source for all two-year condition samples. For the LP method (which used a single year's data) the small sample condition used a random 2/3rds of the year's data. Subsamples differed relative to (1) the inclusion of non-MSU applicants and (2) the size of the calibration pool.

Conditions. A repeated-measures MANOVA design was used (MANOVA<sub>TOM</sub>), using C (methods), A (sample size), N (source of UGPA), and P (predictor intercorrelation) as within-subjects factors. Y (year) was the sole between groups factor. Because all but the four-level N factor was



dichotomous, there were 64 conditions overall. The dichotomous independent variables were, furthermore:

A (sample size), where A= a two-year sample, and B= a three-year sample.

N (source), where M= MSU, and N= all academic origins

P (predictor intercorrelation), where P= uncorrelated predictors, and Q= correlated predictors

Because, in actual practice, sample size variation would be differentially affected by the recruitment approach used, sample size control was imposed by reducing a proportion of usable cases (as opposed to reduction to an absolute numerical limit). Thus from the subsample of cases qualifying for a particular combination of attributes, one condition used the full set of cases, while for a second condition the sample was reduced to (roughly) 2/3 of the full subsample size.

For the conditions using the GC method, the calibrated regression formulas (which used only higher performing applicant cases) were range-restricted relative to the populations to which they were to be applied (namely, the full, annual applicant roster). Because regression coefficients were biased by restriction of the dependent variable, correction for range restriction was appropriate for each standardized partial in the calibrated regression formulas of the general method conditions. Alexander, Carson, Allinger, and Carr (1987) provided the following

formula which was used for correcting doubly truncated correlations (range restriction in both the criterion and in the predictor):

$$\rho = \left[ \frac{\rho'^2 - 1}{2\rho'} \right] \frac{U_x W U_y}{U_x^2 U_y^2 + 1} \pm \left[ \frac{(1 - \rho'^2)^2}{4\rho'^2} \right]^{1/2}$$

Where,

$\rho'$  = range restricted correlation  
 $U_x$  = the ratio of restricted to unrestricted standard deviations for variable X  
 $U_y$  = the ratio of restricted to unrestricted standard deviations for variable Y  
 $\pm$  : corresponds to the sign of  $\rho'$

To correct the GC regression coefficients, a standardized correlation equation must be computed and the partials corrected individually for range restriction.

Dependent Variable. The difference score/dependent variable was a transformed, rank-order difference between rank-of-estimated GGPA and rank-of-actual GGPA (the transformation is referred to as LEWAR: Log of prediction Error -- Weighted, Absolute, and in Rank units) . Because the magnitude of the rank difference was of concern, the absolute value of the difference was used. Where weighting was desired to reflect the proximity of the error to the cut score region, values were weighted to reflect proximity to the lowest GGPA rank. To normalize the distribution of

these absolute values, the modified scores were transformed into their natural logarithms.

Satisfaction of Assumptions. Two critical assumptions for MANOVA are that (1) observations be independent, and that (2) variance be constant (Keeves, 1988). The largely objective nature of admissions and program data allowed some confidence in the proposition that the admissions and program data both had high levels of independence. Differences in condition means were completely determined by the treatment (application of potentially biased, calibrated formulas) and, therefore, there was no treatment-related error to correlate among subgroups. Constant variance was expected due to the lack of differential sampling; a solitary two year sample received all treatments within a repeated-measures format.

Application to Hypotheses. Error variables from these calibration conditions were used to test hypotheses A, B, C, and E.

#### Test of Factor and Year Effects (MANOVA<sub>TOP</sub>)

Sample. Five graduation cohorts (including the 1985 cohort with only a three-year UGPA) were used for the calibration and validity testing of LP models. Random samples were drawn from subsamples of each calibration cohort to obtain calibration subsamples of  $n = 80$  and  $n = 135$ . Subsamples differed relative to (1) the inclusion of

non-MSU undergraduates, (2) the size of the calibration pool, and (3) ability level as measured by PVUGPA.

Conditions. A repeated-measures MANOVA was used (MANOVA<sub>TOF</sub>), including A (sample-size), N (source of UGPA), O (selection on PVUGPA), and P (intercorrelation of predictors) as within-subjects factors. Y (year) was again the sole between-groups factor. Altogether, there were 80 conditions. The dichotomous independent variables were as follows:

A (sample size), where A= 80 random cases of the selected sample, and B= 135 random cases of the selected sample. (For the high-ability condition for 1984, B= 117; and for 1985, B= 111 due to the smaller pools of cases)

N (source), where M= MSU undergraduate origin, and  
N= otherwise

P (predictor intercorrelation), where P= uncorrelated predictors, and Q= correlated predictors

O (achievement level), where G= PVUGPA > 3.0, and  
O= otherwise.

There were five years for the between-subjects variable, year.

Because methods were not being compared in MANOVA<sub>TOF</sub>, sample-size variation was introduced in absolute numbers of cases. Non-MSU cases were qualified as in MANOVA<sub>TOF</sub> (the

methods MANOVA) and predictors were selected as in the methods MANOVA.

Dependent Variable. LEWA-transformed prediction error served as the dependent variable (LEWA= Log of prediction Error: Weighted and in Absolute values). The transformation into logarithms normalized the distribution of the deviations. Prediction-error residuals were weighted according to their proximity to the GGPA rank of one (1).

Application to Hypotheses. Error variables from these calibration conditions were used to test Hypotheses D.

## CHAPTER V

### RESULTS

Table 12 summarizes the research findings relative to the outcomes expected. In a local-proxy (LP) analysis, prediction varied significantly by year, confirming Hypothesis D:

Hypothesis D. With prediction error as the dependent variable, with methods limited to the local-proxy (LP) approach, and with variation controlled with respect to years, sample-size, academic origin, intercorrelation of predictors, and level of ability, prediction will vacillate across years.

Unconfirmed were hypotheses A, B, C, and E. Hypotheses A and B follow:

Hypothesis A. For students falling within a cut score zone, the local-proxy (LP) formula will be more predictive of graduate GPA (GGPA) than will undergraduate GPA (UGPA).

Hypothesis B. For students falling within a cut score zone, the general-mixed (GM) formula will be more predictive of graduate GPA (GGPA) than will the conventional prediction model (the general-criterion formula, GC) as corrected for range restriction.

The failure to find method and year differences contradicted Hypothesis C. Year effect represents differing annual error means in logs of weighted, absolute, prediction-error residuals ( $\alpha = .05$ ). For the following hypothesis, C, prediction (measured by prediction-error means) did not vary across years, although the probability of the observed outcomes was only at .069:

Table 12

## Confirmation of expected outcomes

HYPOTHESIS	CONFIRMATION	
	YES	NO
A. LP more valid than UGPA .....		N
B. GM more valid than GC .....		N
C. Method and Year effects (TOM)		
Method (TOM) .....		N
Years (TOM) .....		N
D. Year effect (TOF) .....	Y	
E. Non-MSU GGPA less validly predicted		N

TOM= Test of Methods

TOF= Test of Factors and Years

LP= Local-proxy estimation formula

UGPA=Undergraduate grade point average

GM= General-mixed estimation formula

GC= General-criterion estimation formula

MSU= Source of UGPA is Michigan State University

GGPA=Graduate grade point average

Hypothesis C. With prediction error as the dependent variable, and with variation controlled with respect to years, methods, sample-size, academic origin, and intercorrelation of predictors, prediction differences among years and methods will be obtained.

For the following hypothesis, E, the numerical outcome was in the right direction, nevertheless, the difference was not significant:

Hypothesis E. Non-MSU undergraduates will be associated with greater prediction error.

#### Tests of Methods (TOM) across Two Years

##### Contrasts Between UGPA and Local-Proxy Estimation

This analysis addressed Hypothesis A: that the LP (local-proxy) approach would out-predict UGPA as an estimator of

GGPA. For the dependent variable, a prediction-error residual transformation was used. The transformation was given the acronym LEWAR, representing Log of prediction Error: Weighted, Absolute, and in Rank units.

The original prediction error residual was the difference in rank between the estimated GGPA and the actual GGPA. It was transformed by the following steps:

- (1) take the absolute value of the difference in ranks between the rank of the estimated GGPA and the rank of the actual GGPA (magnitude and not direction of the error was important),
- (2) weight the absolute value by a non-linear function which is biased towards low GGPA rank (estimation precision is most critical for cases in the cut-score region; for cases which are marginally acceptable),
- (3) obtain the natural log of this weighted error (this serves to normalize the distribution of the variable).

The LEWAR weighting function is the following:

$$\text{LEWAR} = \text{Ln} \left( (1 + \text{ABS}[\text{RKGPA} - \text{RKEST}]) * \left( \frac{[\text{RKGPA} + 1]^3}{\text{RKGPA}^3} + 2 \right) \right)$$

Where LN= a function providing the natural log of a term,

ABS= a function providing the absolute value of a term,

RKGPA= the rank of the GGPA score,

RKEST= the rank of the GGPA estimate

(see Appendix II, Figure IIA for a graph of the weighting used for the LEWAR error transformation).

The LEWAR transformation eliminated the negativity of error residuals, inasmuch as error magnitude (and not



direction of error) was the vital concern. The logarithm transformation was then needed, however, to restore the normal-shaped distribution lost with the transformation of the negative side of the scale. Also included in the transformation was a non-linear weighting (graphed in Figure IIA) which served to increase the importance of prediction-error for cases in the lower range of the GGPA ranking. It was assumed that these cases were most likely to correspond to applicants considered marginal by admission committees. This weighting, which reduced the importance of cases having higher merit rank, served to reduce the actual degrees of freedom to an unknown extent; the systematic deletion of cases is a special case of variation in case weighting. Subsequent significance tests, therefore, must be somewhat liberal.

With respect to assumptions of normality and independence of variables for these conditions, dependent variable frequencies approximated normal distributions and the LEWAR-transformed prediction-error by predictor correlations were mostly non-significant. LEWAR-transformed prediction-error was explicitly and negatively biased according to GGPA rank due to the weighting; however, other significant correlations appeared for variables representing activities and achievements, work experience, narrative writing, MCAT-Chemistry, and average credits carried per term. Only activities and achievements and average credits carried per term ever obtained positive

correlations, and only activities and achievements was consistently positive. GGPA obtained the highest correlation with a LEWAR-transformed variable at  $-.32$  for 1984 data. For 1985, the highest correlations were about ten points less, see the correlation matrix in Appendix III). Except for the intentional bias due to GGPA weighting, there was no reason to expect differential subgroup prediction.

Contrasts using LEWAR-transformed prediction-error as the dependent variable were conducted between UGPA and LP estimation by subtracting the LP discrepancy score (GGPA minus estimated GGPA) from the UGPA discrepancy score.

Table 13 displays four contrasts, not one of which attains statistical significance. Contrasts were conducted separately between 1984 and 1985 data for two (optimum) LP conditions differing only on source of UGPA allowed into the calibration of the formula. The variables contrasted with UGPA were (1)  $n_{MSU}$ ,

where,

- M. only MSU UGPAs were allowed into the calibration,
- B. the calibration used the total year's cases,
- P. intercorrelated predictors ( $>.45$ ) were not permitted into the calibration;

and (2)  $n_{ALL}$ , where,

- N. applicants were not selected on the source of UGPA
- B. the calibration used the total year's cases,
- P. intercorrelated predictors ( $>.45$ ) were not permitted into the calibration

Table 13

## Contrasts of UGPA prediction against LP prediction\*

YEAR	CONTRAST UGPA- LP	DIFFERENCE	tVALUE	DF	PROB
1984	$e_u - e_{nMSU} = 2.7740 - 2.7831 =$	-.0091	-.15	87	.881
1985	$e_u - e_{nMSU} = 2.9655 - 2.9758 =$	-.0103	-.18	100	.855
1984	$e_u - e_{nALL} = 2.7740 - 2.7345 =$	.0395	.59	87	.558
1985	$e_u - e_{nALL} = 2.9655 - 2.9503 =$	.0152	.25	100	.805

\* Dependent variable expressed as LEWAR transformed error=  
 Log of Error: being Weighted, Absolute, and in  
 units of Rank  
 UGPA= Undergraduate grade point average  
 GGPA= Graduate grade point average  
 LP= Mean log of weighted absolute prediction-error in  
 ranks, using local-proxy estimation  
 $e_u$ = Mean log of weighted absolute prediction-error in  
 ranks, using UGPA estimation  
 $e_{nMSU}$ = Mean log of weighted absolute prediction-error in  
 ranks using local-proxy estimation with an all MSU  
 calibration sample  
 $e_{nALL}$ = Mean log of weighted absolute prediction-error in  
 ranks using local-proxy estimation with a  
 calibration sample unselected on UGPA

Not one of the contrasts attained significance, thus ruling out the possibility of significance familywise. The test of Hypothesis A, therefore, did not confirm greater validity for the LP estimate over UGPA for cases in the cut-score region.

Comparisons Against the General-criterion Method.

These analyses addressed (1) Hypothesis B: that the GM (general-mixed) approach would out-predict the GC (general-criterion), and (2) Hypothesis C: that prediction would differ by method used and by year's data estimated.

Methods of selection formula calibration were evaluated with respect to mean level of LEWAR (the dependent variable). Thus significant effects (MANOVA/ ANOVA) would indicate less valid estimation. GC equations were calibrated using standardized regression in order to allow corrections of partials (standardized regression coefficients) for restrictions-of-range. The proxy method equations used raw-score regression. Precision between the two types of method equations was compared in terms of LEWAR-transformed residuals of rank (between estimated GGPAs and actual veterinary school GGPAs). Method difference in estimation across the whole range of GGPAs is of some interest, although differences in estimation would be most relevant to the cases close to the cut-score. For instance, the college's selection ratio may expand, invalidating the previous selection procedures for a particular cut-score. As a general comparison of methods, Table 14 presents average predictive rank among the four calibration methods where error is reported unweighted in its natural log (log of absolute rank error). No differences were significant.

Table 15 presents between-subject by within-subject effects for the test-of-methods using a repeated-measures MANOVA (MANOVA<sub>TOM</sub>) with the following within subject factors:

Table 14

Method efficiency in logs of absolute rank-error\*

	METHOD LOG OF ABSOLUTE RANK ERROR			
	GC	LP	GP	GM
MEAN	2.50	2.43	2.71	2.09
s.d.	1.31	.71	1.10	1.08
s.e.	.34	.18	.28	.28

\* Rank-error is absolute value of a discrepancy: estimate rank minus rank of GGPA

GC= General-criterion,

LP= Local-proxy

GP= General-proxy,

GM= General-mixed (authentic and proxy criterion)

C= calibration methods,

Q= intercorrelation of calibration predictors,

N= source of calibration applicants, and

A= size of the calibration sample.

Years (Y) were the only between subject factors.

The dependent variables were prediction rank-error residuals for a factorial matrix of 64 calibration conditions. The average absolute prediction-error in ranks across conditions ranged from 17 to 34 positions. These residuals received the LEWAR transformation (described on page 78, 79). The plots of the dependent variable frequencies approximated normal distributions and the predictor by dependent variable correlations were often marginally significant for predictors similar to GGPA (due to explicit weighting). As presented in Table 15, the a priori contrast between the GM (general-mixed) and GC (general-criterion) conditions was not significant, indicating comparable prediction. Table 16 reports no

Table 15

Estimate of average contrast between general-criterion and  
general-mixed transformed prediction-error means

SOURCE OF VARIATION	COEFF	ST.ER	T-VALUE	SIG.T	L-BND	H-BND
GM minus GC	-.092	.107	-.863	.389	-.304	.119
* Measured in transformed prediction-error residuals (LEWAR= Log Error: Weighted, Absolute, and in Rank units)						
GM= general mixed: multi-year calibration sample with authentic and proxy criterion						
GC= general criterion: multi-year calibration sample with authentic criterion						

differences among methods; Table 17 likewise reports no  
differences among years (although for years, the  
probability of the results under the null hypothesis is  
only .069).

Table 16

Test of main effect for methods\* (MANOVA<sub>TOM</sub>)

M U L T I V A R I A T E					
SOURCE-OF-VARIATION	WILKS-LAMBDA	MULT-F	HYP-DF	ERR-DF	SIG
Methods	.99160	.52258	3	185	.667
* Measured in transformed prediction-error residuals (LEWAR= Log Error: Weighted, Absolute, and in Rank units)					
Methods= local-proxy, general-criterion, general-proxy, & general-mixed					
MANOVA <sub>TOM</sub> = Test-of-methods repeated-measures MANOVA					

Table 17

Test of main effect for years\* (MANOVA<sub>TOM</sub>)

SOURCE OF VARIATION	SS	DF	MS	F	PROB
Year	66.09	1	66.09	3.34	.069

\* Measured in transformed prediction-error residuals  
 (LEWAR=Log Error: Weighted, Absolute, and in Rank units)  
 MANOVA<sub>TOM</sub>=Test-of-methods repeated-measures MANOVA  
 Years=1984, 1985

Plots of the means in Figures 1 and 2 (for 1984 and 1985 cohorts respectively) illustrate the contrasting interactiveness between the GC (general-criterion) and the GM (general-mixed) methods. Under the following conditions the GC method is notably less predictive: (1) the sample is small, (2) the calibration sample is highly selected on a single source of applicants (relative to the graduating cohort), and (3) the calibration does not screen out correlated predictors (note condition AMQ for 1984 and 1985; where,

A= sample size= small  
 M= source of UGPA= MSU  
 Q= intercorrelation of predictors > .45).

It should be noted that the GC method uses a much smaller calibration sample than do the competing proxy methods, although all small-sample conditions use two-cohort calibration samples. For the GC method, this amounted to about 200 applicants, for the proxy methods, over 400 applicants were available per the two cohorts.

Table 18

## Cross-validities for prediction methods

METHOD	CONDITION	COHORT 1984 n= 87	COHORT 1985 n= 101
General Criterion . . . .	NQ . . . .	.60 . . . . .	.49
General Mixed . . . .	NQ . . . .	.58 . . . . .	.34
General Proxy . . . .	NP . . . .	.54 . . . . .	.38
Local Proxy . . . .	MP . . . .	.53 . . . . .	.40
PVUGPA . . . . .		.49 . . . . .	.37
UGPA . . . . .		.55 . . . . .	.41

Method condition acronyms represent the following:

N= source= . . . . . all UGPAs

M= source= . . . . . MSU UGPA

Q= intercorrelation> . . . . .45

P= intercorrelation< . . . . .45

Inclusion of (1) UGPA error in the calibration sample (a consequence of multiple UGPA sources in the calibration sample) and (2) the reduction of correlated predictors, appears to reduce the adverse impact of small calibration sample size (note condition ANP for 1984 and 1985; where,

A= sample size= small

N= source of UGPA= all

P= intercorrelation of predictors < .45).

Table 18 presents cross-validity coefficients for the presumed optimal condition for each of the four methods when applied to predict 1984 and 1985 GGPA from the appropriate admissions data. It is apparent from the table that, numerically, proxy criterion GGPA estimates (from all but the general criterion method) correlate somewhat below the correlation level of conventional (GC) multiple predictor GGPA estimates (from the general criterion



method). Because cross-validation was performed on a restricted sample (applicants with a subsequent GGPA), however, it cannot precisely index method validity for application to ordinary admissions selection (where the range of the criterion is unrestricted). To compensate for this problem, the regression weights for the GC (general-criterion) were corrected for restriction prior to the computation of the cross validity coefficients. This adjusted the GC range-restricted weights towards weight levels appropriate for a wide-range criterion. Problems with using correlation coefficients to evaluate selection validity are discussed later.

The test-of-methods addressed Hypotheses B and C as follows:

- B. Hypothesis B, that GM (general-mixed) prediction would exceed that of the GC (general-criterion) approach, was not confirmed. Conversely, under the most normal conditions (e.g. larger calibration sample-size, and multiple source UGPAs, the GC method generally appeared more predictive than other methods for cases in the cut-score region,
- C. Hypothesis C, that main prediction differences among methods and years would emerge was not confirmed, suggesting that these factors were unimportant to cases in the cut-score region. Nevertheless, the probability of the outcome for years (under the null hypothesis) was low ( $p = .07$ ).

# CONDITION MEANS (MANOVAtom) 1984

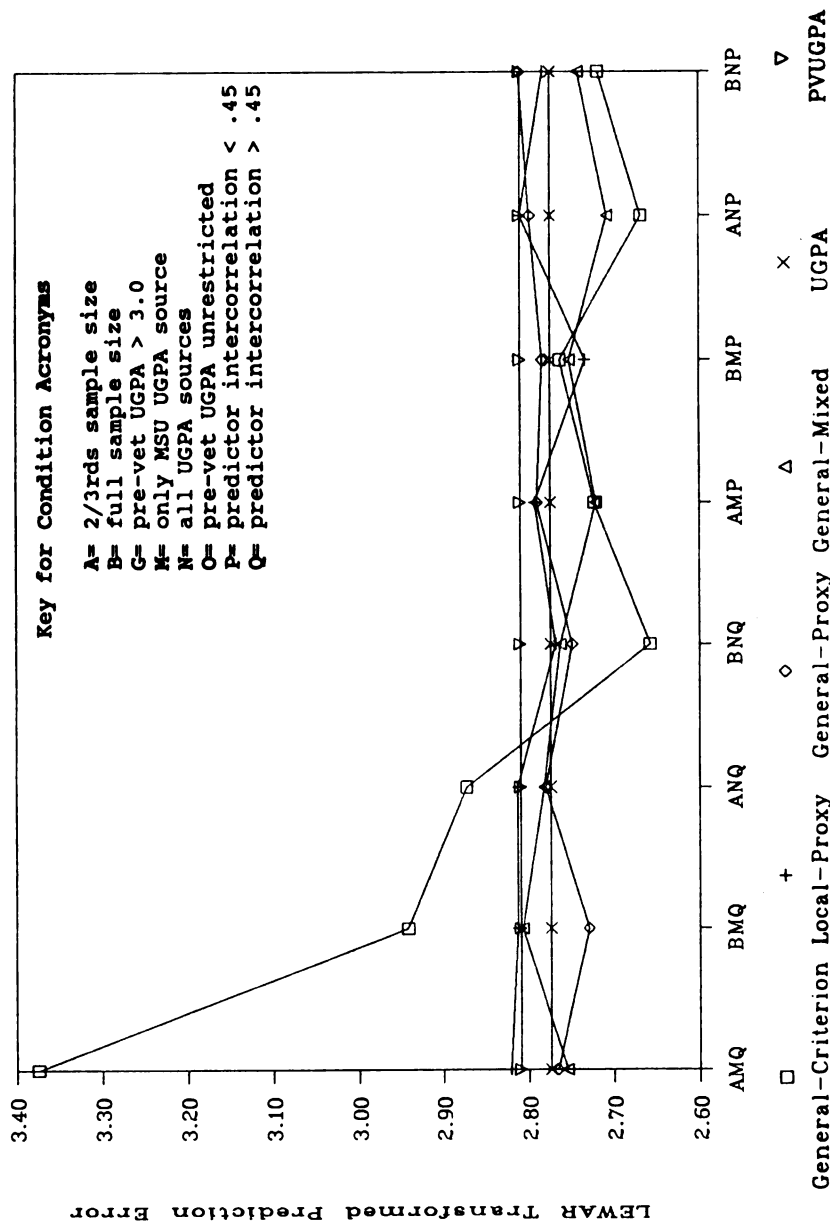


FIGURE 1. Relative levels of transformed prediction error among GGPA estimation methods (1984)

# CONDITION MEANS (MANOVAtom) 1985

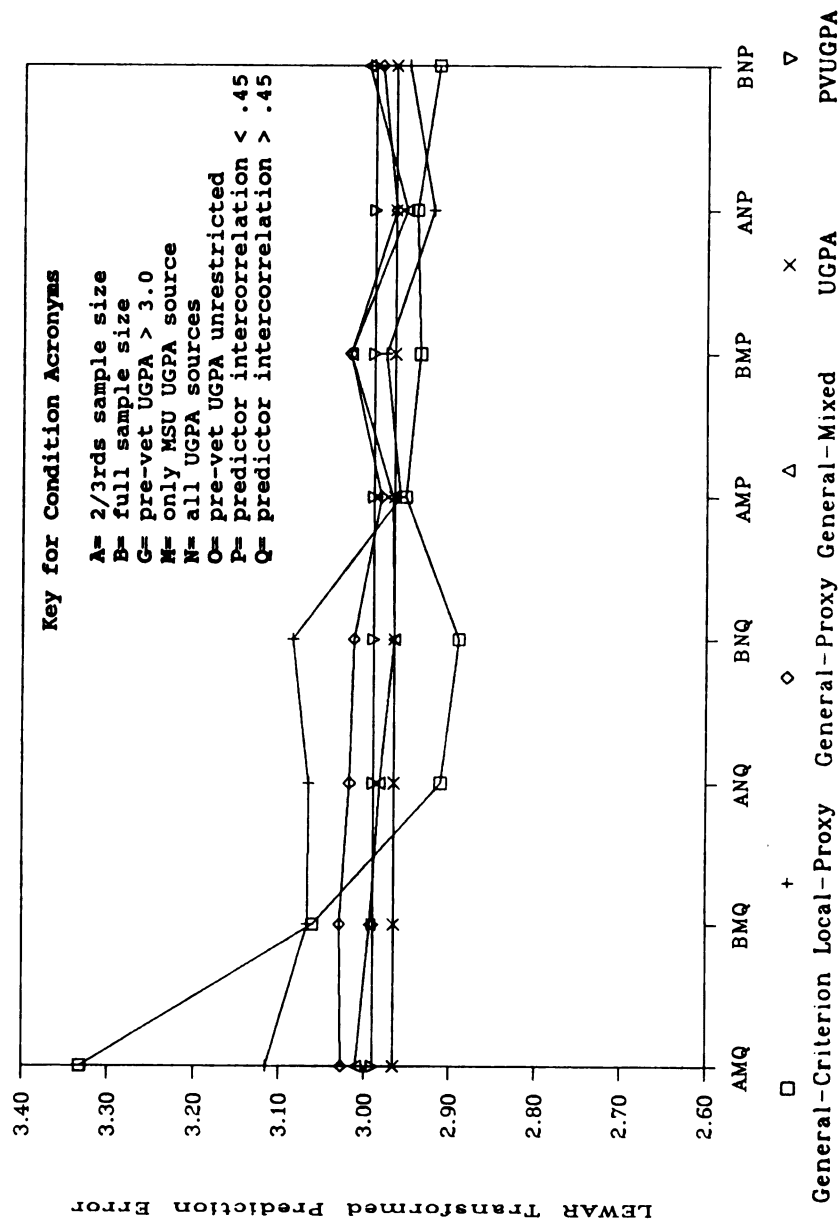


FIGURE 2. Relative levels of transformed prediction error among GGPA estimation methods (1985)

### Test of Years and Other Factors under LP (local-proxy)

Independent Variable Contrasts. This analysis addresses Hypothesis D which predicts an LP method prediction difference among years. The dependent variable was a transformed prediction-error residual differing from LEWAR variables by its use of actual rather than rank error (see pages 78 and 79). Its transformation is referred to as LEWA, representing Log Error: Weighted and Absolute. The use of actual prediction error (rather than the predictive ranking error used in the LEWAR-transformed dependent variable) allowed more powerful significance testing than in the test of methods analysis.

The original prediction-error residual was the difference in the estimated GGPA and the actual GGPA. Across conditions, the average absolute prediction-error ranged from .29 to 3.87. The LEWA transformation is as follows:

- (1) take the absolute value of the difference between  
the estimated GGPA and the actual GGPA (magnitude  
but not direction of the difference is of interest)
- (2) weight the absolute value by a non-linear function  
which is biased towards low GGPA rank (see Appendix  
II for a graph of the weighting function used to  
give greater weight to the cut-score region),
- (3) obtain the natural log of this weighted error (to  
normalize the distribution of the error variable).

The LEWA weighting function is the following:

$$\text{LEWA} = \text{Ln} \{ (1 + \text{ABS} [\text{GGPA} - \text{ESTGGPA}]) * \left( \frac{[\text{RKGPA} + 1]^3}{\text{RKGPA}^3} + 2 \right) \}$$

Where,

LN= a function providing the natural log of a term,  
 ABS= a function providing the absolute value of a term,  
 RKGPA= the rank of the GGPA score,  
 ESTGGPA= the GGPA estimate.

LEWA-transformed variable distributions were approximately normal. Due to deliberate weighting by GGPA rank, LEWA variables correlated negatively with predictors which were similar to GGPA, and positively with predictors which were inversely related to GGPA. Negative correlations with GGPA were as large as  $-.75$ . (Correlations were as great as  $.66$  for age,  $.70$  for possessing a college degree, and  $-.45$  for average credits taken per term. Generally, however, correlations were not significant).

A test-of-factors-and-years MANOVA ( $\text{MANOVA}_{\text{TOF}}$ ) was performed with five levels for the between variable, year. For  $\text{MANOVA}_{\text{TOF}}$ , it was possible to include an additional calibration factor, range restriction, defined as restricted where the PVUGPAs (undergraduate prerequisite veterinary course UGPAs) were above 3.0 and not restricted otherwise. Table 19 displays the ANOVA table confirming prediction differences among years among these conditions (and thereby confirming Hypothesis D). For cases in the

Table 19

Test of main effects for years\* (MANOVA<sub>TOF</sub>)

SOURCE OF VARIATION	SS	DF	MS	F	PROB
Within Cells	643.78	462	1.39		
Constant	1619.13	1	1619.13	1161.95	<.001
Years	33.60	4	8.40	6.03	<.001

\* Measured in LEWAR-transformed prediction-error residuals  
 LEWAR=Log Error: Weighted, Absolute, and in Rank units)  
 MANOVA<sub>TOF</sub>=Test-of-factors & years a repeated-measures  
 MANOVA  
 Years= 1981, 1982, 1983, 1984, 1985

cut-score region, LP (local-proxy) estimation differed across years.

To further amplify the interaction among calibration factors, Figures 3 through 8 graph the dependent variable means (prediction error) for selected contrasting conditions. In Figure 3, the better predictiveness of the non-restricted condition is weakly suggested by the relative levels of error. (compare unrestricted ANOP,

[where A= sample size= small  
 N= source of UGPA= all  
 O= range restriction= unrestricted  
 P= intercorrelation < .45]

against restricted ANGP,

[where A= sample size= small  
 N= source of UGPA= all  
 G= range restriction= on  
 P= intercorrelation < .45])

# RANGE EFFECT OVER FIVE COHORTS

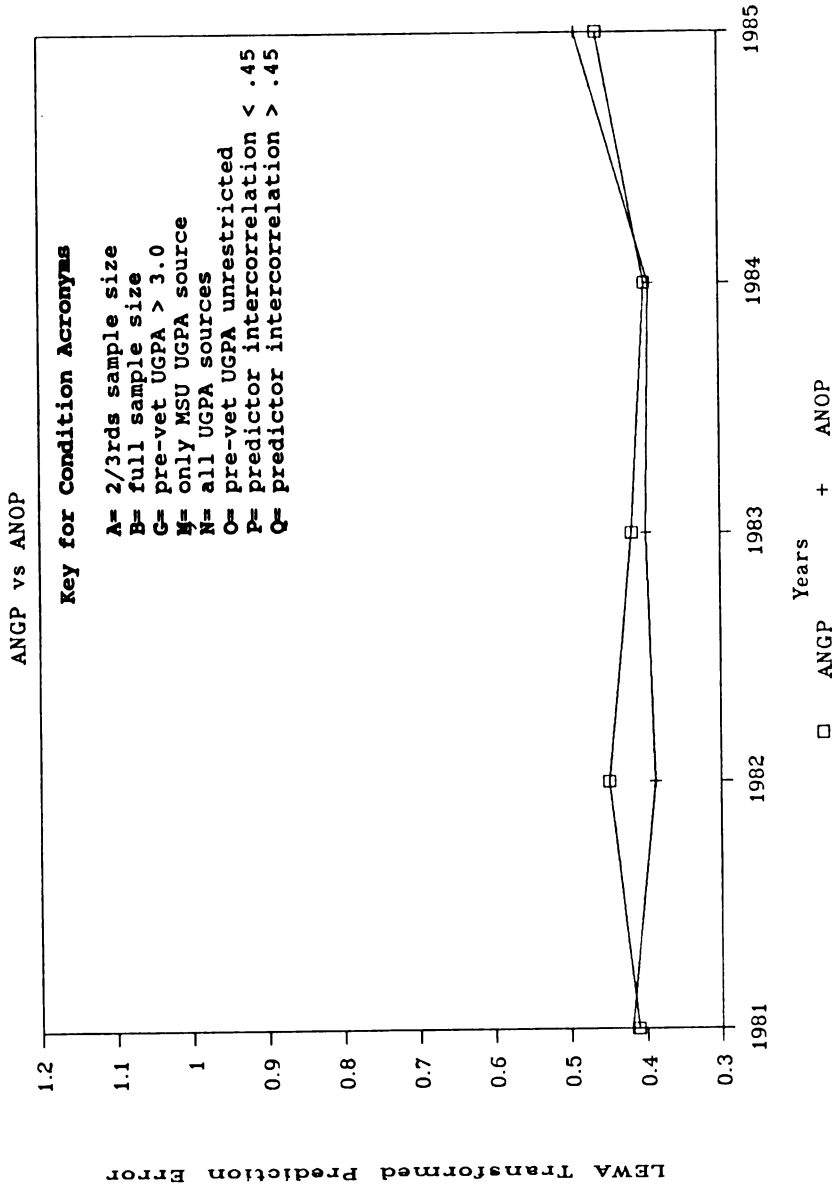


FIGURE 3. Prediction error under local-proxy prediction where the proxy-criterion has been restricted

# INTERCORRELATION EFFECT: OVER COHORTS

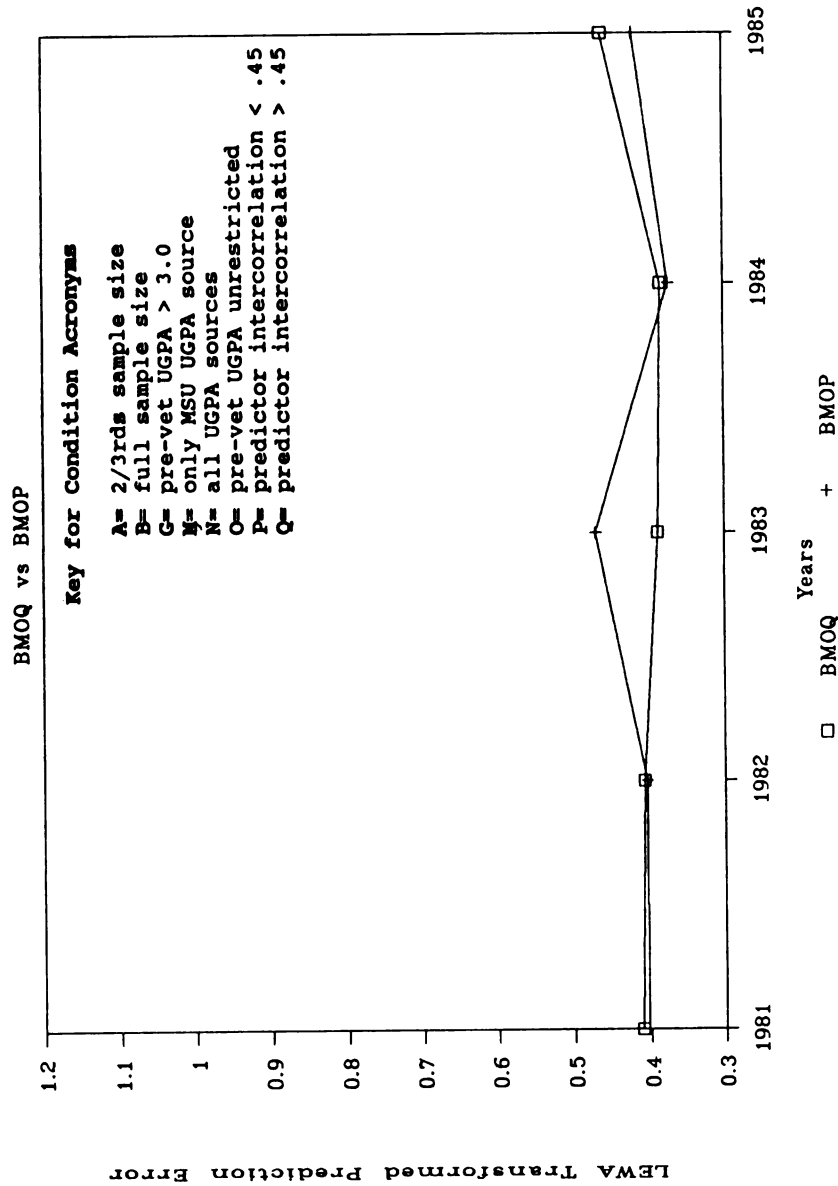


FIGURE 4. When measured by transformed prediction error, local-proxy prediction appears little affected by level of intercorrelation



In Figure 4, the difference in efficiency between calibration conditions BMOQ

(where B= sample size= large  
M= source of UGPA= MSU  
O= range restriction= off  
Q= intercorrelation > .45)

and BMOP,

(where B= sample size= large  
M= source of UGPA= MSU  
O= range restriction= off  
P= intercorrelation < .45;

differing on intercorrelation of predictors; BMOQ having greater intercorrelation) fails to indicate a consistent pattern of difference. In Figure 5, BMOQ and BMGQ

(where B= sample size= large  
M= source of UGPA= MSU  
G= range restriction= on  
Q= intercorrelation > .45;

differing by restriction of range; BMGQ being restricted), differ dramatically in 1984-- a dramatic interaction for years by range restriction. In Figure 6, BMOQ and BNOQ

(where B= sample size= large  
N= source of UGPA= all  
O= range restriction= off  
Q= intercorrelation > .45;

differing by the inclusion of non-MSU undergraduates; BNOQ being the all-sources condition) differ substantially in both 1984 and in 1985, an interaction this time between years and non-selected UGPA source (in computing the formula calibration).

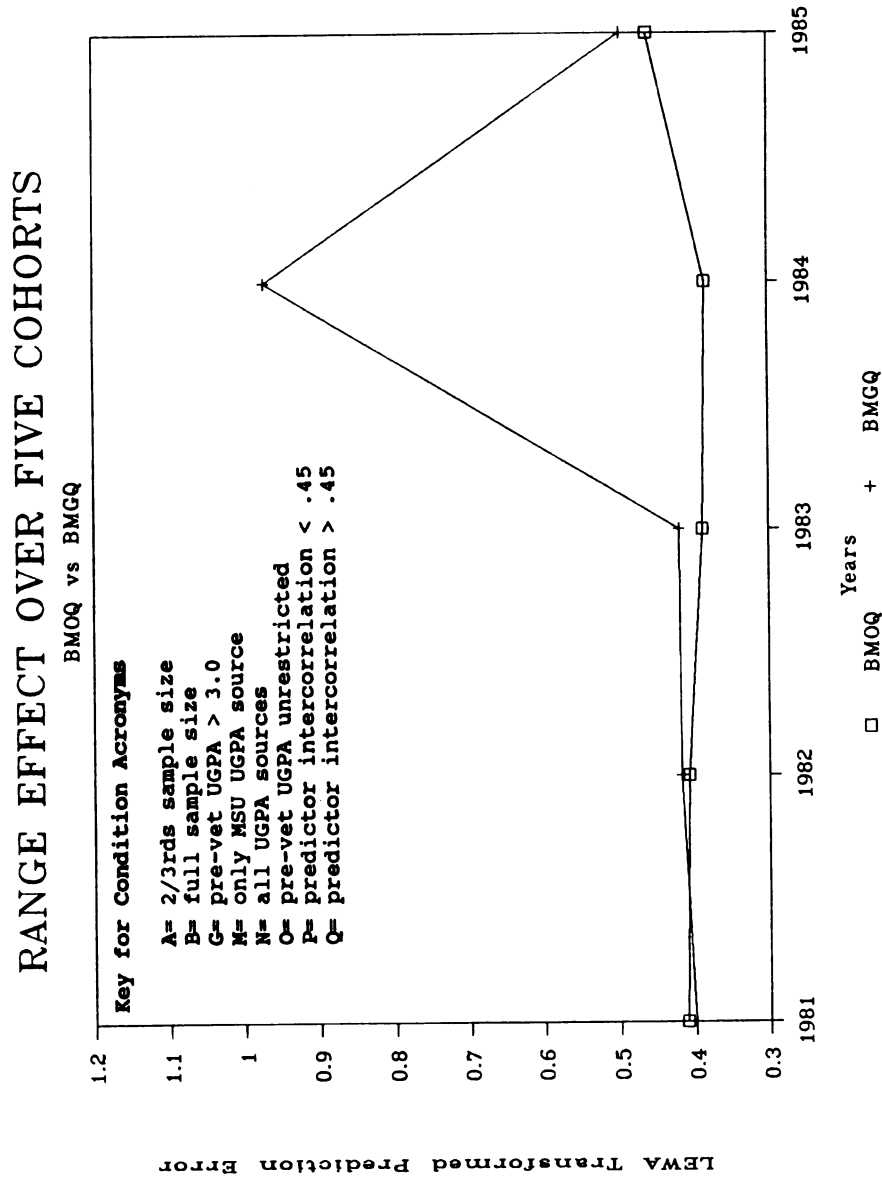


FIGURE 5. When measured by transformed prediction error, restriction of the proxy-criterion range only appears to affect prediction for the 1984 cohort

# UGPA-SOURCE: IMPACT OVER COHORTS

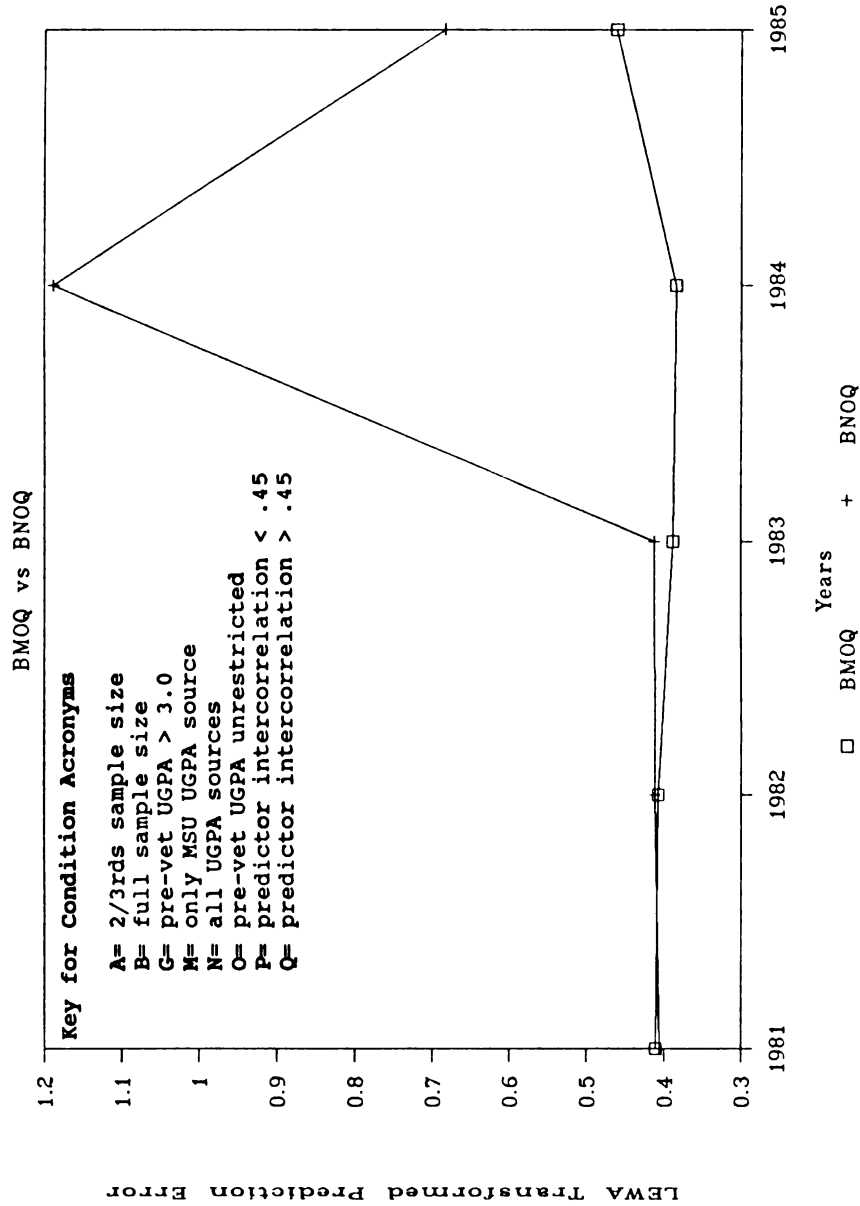


FIGURE 6. When measured by transformed prediction error, source of UGPA only appears to affect prediction for the 1984 and 1985 cohorts

# INTERCORRELATION EFFECT: OVER COHORTS

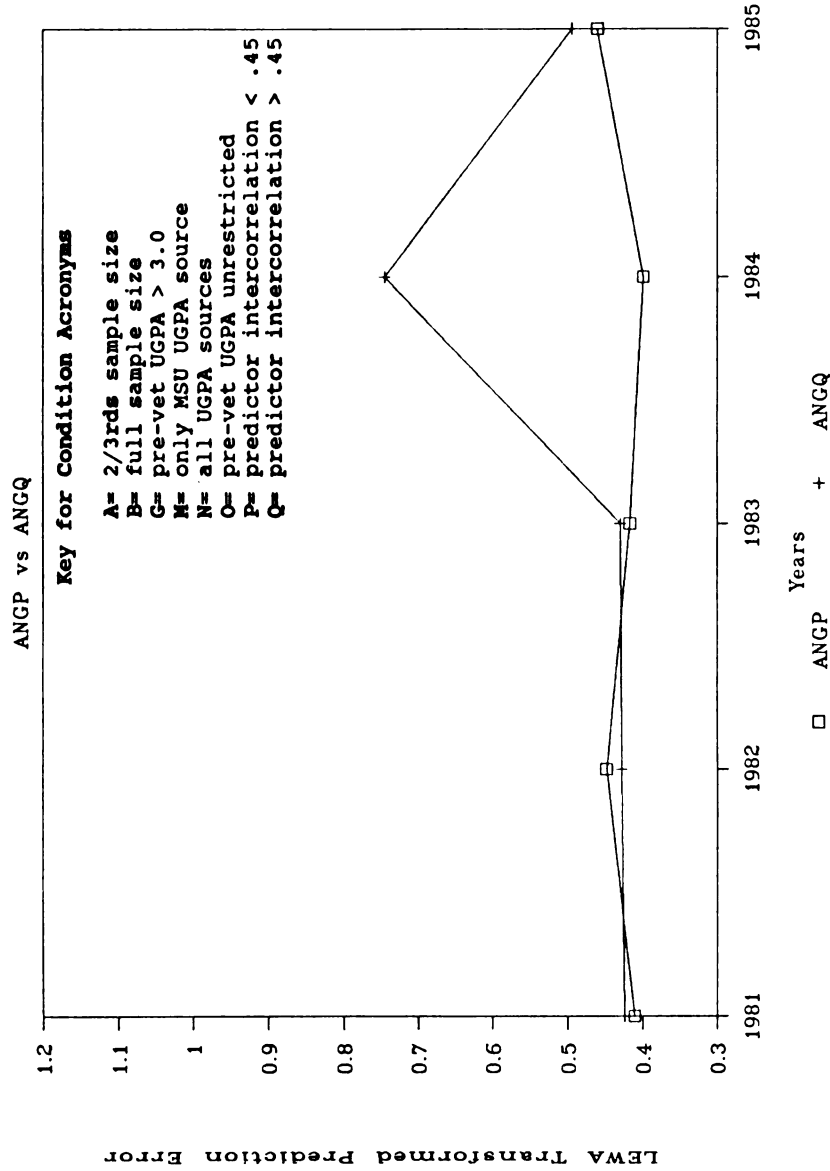


FIGURE 7. When measured by transformed prediction error, local-proxy prediction appears affected by level of intercorrelation for the 1984 cohort

# FULL VS 2/3RDS SAMPLE SIZE

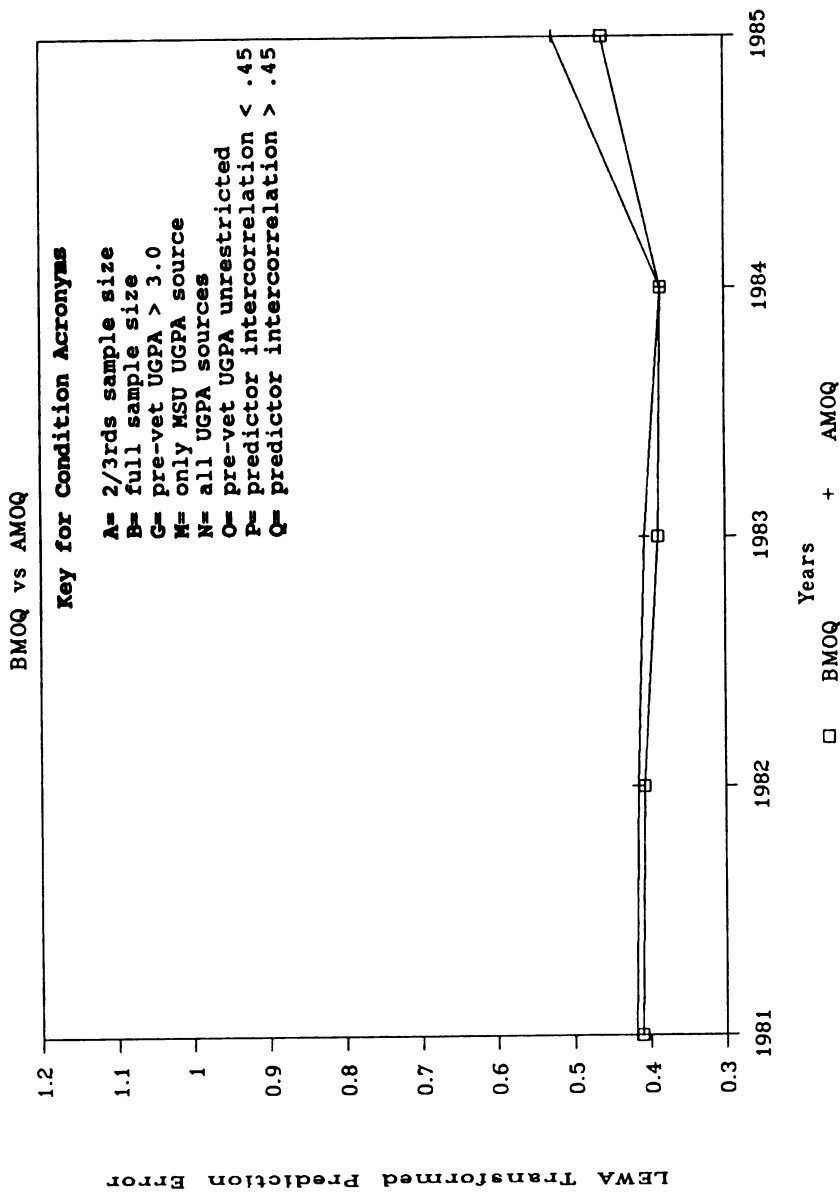


FIGURE 8. When measured by transformed prediction error, sample size appears to have little effect on prediction

In Figure 7, the ANGP-ANGQ contrast

(where A= sample size= small  
 N= source of UGPA= all  
 G= range restriction= off  
 Q= intercorrelation of predictors < .45;

differing by intercorrelation of predictors) also interacts with year in 1984. In Figure 8, a difference in sample size between BMOQ and AMOQ

(where A= sample size= small  
 M= source of UGPA= MSU  
 O= restriction of range= unrestricted  
 Q= intercorrelation of predictors < .45)

results in small differences in prediction error until 1985 when year interacts with sample size.

This graphical evidence is most noteworthy for its illumination of the interactive effect of years on all of the factors: (N) source, (A) sample size, (G) restriction of achievement level, and (Q) intercorrelation of predictors. It may be recalled that in the calibration for this test of years and factors, the large sample size was  $n = 135$ , while the small size was  $n = 80$ ). It should be noted, therefore, that in the case of intercorrelated predictors ( $r > .45$ ), the interactive effect of years only occurs where the sample is small (compare ANGP:ANGQ in Figure 7 with BMOP:BMOQ in Figure 4). Surprisingly, an interactive pattern appears under the large sample, restriction of PVUGPA, which doesn't occur under the small sample condition (in Figure 5, contrast BMOQ:BMGQ, then

compare with ANGP:ANOP in Figure 3). It happened that for the years 1984 and 1985, the large sample, restricted PVUGPA conditions had low case counts (1984:  $n = 117$ , 1985:  $n = 111$ ). Due, perhaps, to the greater sampling error for these two years, the selection for these large sample conditions wasn't representative (and was, perhaps, greatly restricted for MSU applicants). By chance, the small sample conditions for these two years select a much more representative sample.

#### Relative Validity of Non-MSU and MSU UGPAs

Contrary to Hypothesis E, Table 20 shows that non-MSU applicants were not associated with significantly greater prediction error (in 1981 and 1982). For these years, log-transformed absolute ranking error does appear to be

Table 20

T-test of mean error\* between MSU and non-MSU applicants

ERROR TYPE	N	MEAN	SD	SE	T	DF	PROB(2-TAIL)
<b>LEAR GC<sub>Cg</sub></b>							
N-MSU	26	2.7924	1.045	.205	.76	187	.449
MSU	163	2.6307	1.003	.079			
<b>LEAR UGPA</b>							
N-MSU	26	2.9755	.866	.170	1.42	187	.189
MSU	163	2.7108	.963	.075			

\* Measured in transformed, unweighted prediction error residuals (log of absolute rank error)

GC<sub>Cg</sub> = General-criterion formula where (1) sample size is large, (2) intercorrelated predictors are used, and (3) non-MSU UGPAs are included

slightly greater when the estimator of GGPA is UGPA as opposed to the GC (general-criterion) estimator.



## CHAPTER VI

### DISCUSSION

This study enquires whether prediction (or estimation) using a proxy criterion might exhibit better precision than conventional methods (use of UGPA, and use of regression estimates from multiple predictors of previous annual samples). It is assumed that moderation of prediction error by years is a data characteristic favoring the practice of local-proxy estimation. An ancillary concern is the role of observable prediction factors which might influence the relative validity between proxy and authentic criterion prediction methods.

The experimental proxy criterion methods do not significantly differ from the conventional methods when evaluated on the basis of logs of weighted absolute values of prediction error residuals (weighted according to proximity to the cut-score). Thus, for these data and methods, method of estimation seems to have little impact on cases most vulnerable to rejection. Nevertheless, no predictive advantage is observed for any of the proxy criterion methods over that of the simple UGPA (moreover, by the same standard, Figures 1 and 2 suggest no significant advantage over UGPA for the conventional multiple predictor method either).

For the study and control of prediction factors (Q= level of intercorrelation of predictors, N= source of UGPA, and A= sample size) in the methods analysis (or MANOVA<sub>TOM</sub>), two levels for each factor were represented in the condition matrix. In addition, a between groups factor, years, also had two categories, 1984 and 1985. For these years, factors did not significantly differ from each other in their influence on prediction.

Likewise for the test-of-factors-and-years (or MANOVA<sub>TOF</sub>), similar prediction factors were in the condition matrix plus another two-level factor, range restriction (O) on the veterinary pre-requisites UGPA. The methods factor was eliminated; all conditions used the LP (local-proxy) approach. The sole between groups factor, years, had five categories (five years of data). For these years, three of four factors differed from each other in their influence on prediction. Factor scores for factors predictor intercorrelation (Q), and for restriction-of-range on the veterinary prerequisites UGPA (O), differed from each other and from the factor scores for sample size (A) and applicant source (N) factors. The latter factor scores, however, failed to differ significantly from each other. While this study does not find a consistent predictive advantage for the proxy methods over the GC (general-criterion method) or even over the UGPA, (1) some potential enhancements of the method remain untried, and (2) the theoretical viability of the concept is not

clearly refuted. Nevertheless, from Figures 1 and 2, it appears that the UGPA acting alone is virtually as valid a predictor of GGPA as is multiple variable prediction (particularly when (1) ranking-error is the error variable, and where (2) cut-score region cases are emphasized).

#### High UGPA Validity

Standardized graduate admissions examinations provide consistent scales of evaluation across schools, and may be indirectly responsible for the apparent reliability of UGPA that was present in the data studied; local grading policies may be shaped by national admissions test performance. The same may also be true for other admissions variables; effective variables may provide little independent prediction where their effects are mediated through the UGPA. It is probably more likely the case, however, that the sounder psychometric indicators (the UGPA and the admissions tests) are more valid as predictors due to (1) superior reliability and due to (2) a factor structure similar to the GGPA.

The adequacy of UGPA prediction can be expected to diminish, however, where a larger proportion of multiple-source UGPAs have to be handled. This would be the case in prediction across sites and in non-veterinary graduate programs which may be less influenced by geography (veterinary programs serve primary geographic regions and are, on average, less than one per state). Likewise, GC

(general-criterion) prediction would be less precise due to measurement error in the independent variables during formula calibration, and due to the subsequent lack of precision in appropriately applying formula weights to larger numbers of ambiguous UGPA scores. The localizing of prediction limits the entry of uncontrolled moderators into the statistical analysis, potentially improving internal validity, but it nevertheless diminishes external validity. The present study is local with respect to the dependent variable (100% MSU GGPA), and it happens to be mostly local with respect to the UGPA predictor. The generalizable prediction methods in this study (GC, GP, and GM) are largely limited to generalizing across years. For this data there is only weak evidence that years moderate prediction (plots of means indicated only one of five years where prediction error differed notably). Moderation-due-to-years in other sites, and moderation-by-sites remain to be investigated.

#### Potential Usefulness of Proxy Methods

Although it is true that problems due to error-laden UGPAs also impair prediction (and selection) with proxy-criterion methods, the potential exists for situations where proxy methods may be optimal. One such occasion may exist where changes are being made in the outcome variable.

Use of the GGPA as a prediction criterion is questionable because it can fail to adequately represent

practical competence. Ultimately, however, the problem is not the limitations of GGPA, per se, but of the measurement design and procedures which succeed or fail to measure appropriate performance factors. Typically, this measurement issue is ignored while attention is fixed on the issue of selecting and weighting adequate predictors. Researchers in this field often conclude that the prospects of predicting beyond a given level of precision may be futile. Such a conclusion, however, neglects to address the dependence of prediction on the quality of measurement which determines the criterion variable. The measurement of factors which distinguish adequately between levels of professional competence must be the most important component to the improvement of academic selection. As institutions depart, nevertheless, from measurement conventions (across years or locations) the need for localized prediction will emerge. It may be in such a context that proxy-criterion prediction methods find greater practicality.

#### Potential Improvements to Estimation with a Proxy Criterion

Figures 1 and 2 indicate that for the two years represented, the GC method appears to consistently out-predict UGPA when the calibration condition is BNQ

(where      B= sample size= large  
              N= source of UGPA= all  
              Q= predictor intercorrelation >.45),

**AMP**

(where A= sample size= small  
M= source of UGPA= MSU  
P= predictor intercorrelation  $<.45$ ),

**BMP**

(where B= sample size= large  
M= source of UGPA= MSU  
P= predictor intercorrelation  $<.45$ ),

**ANP**

(where A= sample size= small  
N= source of UGPA= all  
P= predictor intercorrelation  $<.45$ ),

**or BNP**

(where B= sample size= large  
N= source of UGPA= all  
P= predictor intercorrelation  $<.45$ );

the LP method predicts similar to the UGPA condition for the two years under the calibration condition BMP; and the GM method is consistently as predictive as UGPA under calibration conditions BNQ, ANQ

(where A= sample size= small  
N= source of UGPA= all  
Q= predictor intercorrelation  $>.45$ ),

and ANP. Hypotheses are hereby provided to account for these patterns.

In Table 21, factors are specified (IDEAL FACTORS) under each method-sample size combination. Each ideal factor is based on a logical or empirical expectation relating (1) intercorrelation of predictors ( $Q= r > .45$ ,  $P=r < .45$ ) to sample size (large  $> 270$ , small= 180), (2) source of UGPA (N= all sources, M= MSU) to type of criterion used, and (3) source of UGPA to type of UGPA predictor used. For method GC with a small sample size, P

Table 21

A chart for identifying ideal method prediction factors

METHOD	APPLICATION		CALIBRATION		
	Variable Estimated	Predictor Used	N	Criterion Used	Predictor Used
GC-small smple GGPA <sub>4</sub> IDEAL FACTORS . . . . .		UGPA	180	GGPA <sub>x</sub> <P>	UGPA <sub>x</sub> <N>
GC-large smple GGPA <sub>4</sub> IDEAL FACTORS . . . . .		UGPA <sub>4</sub>	270	GGPA <sub>x</sub> <Q>	UGPA <sub>x</sub> <N>
LP-small smple GGPA <sub>4</sub> IDEAL FACTORS . . . . .		UGPA <sub>4</sub>	200	PVS <sub>4</sub> <P> <M>	UGPA <sub>4</sub> <N>
LP-large smple GGPA <sub>4</sub> IDEAL FACTORS . . . . .		UGPA <sub>4</sub>	300	PVS <sub>4</sub> <M>	UGPA <sub>4</sub> <N>
GP-small smple GGPA <sub>4</sub> IDEAL FACTORS . . . . .		UGPA <sub>4</sub>	600	PVS <sub>x</sub> <P> <M>	UGPA <sub>x</sub> <N>
GP-large smple GGPA <sub>4</sub> IDEAL FACTORS . . . . .		UGPA <sub>4</sub>	900	PVS <sub>x</sub> <P> <M>	UGPA <sub>x</sub> <N>
GM-small smple GGPA <sub>4</sub> IDEAL FACTORS . . . . .		UGPA <sub>4</sub>	600	GGPA <sub>x</sub> PVS <sub>x</sub> <P>	UGPA <sub>x</sub> UGPA <sub>x</sub> <N>
GM-large smple GGPA <sub>4</sub> IDEAL FACTORS . . . . .		UGPA <sub>4</sub>	900	GGPA <sub>x</sub> PVS <sub>x</sub> <Q>	UGPA <sub>x</sub> UGPA <sub>x</sub> <N>

GGPA<sub>4</sub> = estimated graduate GPA of applicantUGPA<sub>4</sub> = applicant's undergraduate GPAPVS<sub>4</sub> = applicant's veterinary prerequisites UGPAGGPA<sub>x</sub> = includes GGPAs of applicants of other yearsUGPA<sub>x</sub> = includes UGPAs of applicants of other yearsPVS<sub>x</sub> = includes PVS<sub>4</sub>s of applicants of other years

&lt;P&gt; = predictor intercorrelation &lt; .45

&lt;Q&gt; = predictor intercorrelation &gt; .45

&lt;N&gt; = UGPA not selected by source

&lt;M&gt; = only MSU UGPAs in sample

(predictor intercorrelation  $< .45$ ) is recommended due to the small sample size, while N (a calibration sample containing some non-MSU applicants) is recommended due to the fact that the  $UGPA_4$  (UGPA for the fourth year cohort) used in the application stage will be error-laden relative to the frequency of non-MSU applicant UGPAs (if factor M were used, only MSU UGPAs would be in the calibration, thus potentially inflating the beta weight for the UGPA predictor). For all conditions, N is generally appropriate because for all of the applicant data there are non-MSU UGPAs which impose scaling error into the data. Nevertheless, for methods LP and GP, M (specifying a calibration sample containing only MSU applicants) provides a calibration criterion having less error than one which includes non-MSU applicants. Because it is impossible to use UGPA from all sources as a predictor and to simultaneously use a PVUGPA (proxy criterion) from only MSU cases, a choice must be made between two advantages: (1) a realistic predictor or (2) a less fallible proxy criterion. The advantage from M (a less fallible proxy criterion) is arbitrarily granted more importance than that from N and thus M is recommended for LP and GP.

Where the sample size is sufficiently large and the calibration criterion adequately precise, Q (predictor intercorrelation  $> .45$ ) seems to improve prediction, otherwise P (predictor intercorrelation  $< .45$ ) seems more efficient.



Having used the present data to inform this set of expectancy rules, the hit rate for these rules on the same two years of data is 75%. The factor that is difficult to specify with confidence is  $Q$  (predictor intercorrelation) because (1) the calibration criterion is imprecise and because (2) the sample size is only moderately large.

Study of the condition means, therefore, gives rise to three hypotheses: (1) LP prediction may be improved by the use of single-source selected cases (to improve the reliability of the proxy-criterion when calibrating the selection formula), (2) use of an overly reliable UGPA predictor (in the calibration) may contribute to prediction error; hence, the addition of random error to the UGPA predictor in the calibration may reduce the inflation of the UGPA predictor weight, and (3) multicollinearity may be less of a problem with these variables than has been presumed; given an adequate sample size and precise measures, moderately correlated multiple prediction may be more valid than prediction with correlated variables selected out.

#### How Intercorrelation May Remain Benign

As Pedhazur (1982) points out, there is no agreement on the breadth of meaning in the term multicollinearity, although its existence is unambiguous where predictor intercorrelation biases regression coefficients. It is clear that parameter bias is more likely as (1) the number

of predictors approaches the number of cases in the calibration sample, (2) the predictors share mutual factors, (3) intercorrelated predictors are highly correlated with the criterion, and (4) predictors lack reliability (Kenny, 1982). Although for this study GC (general-criterion) performance is limited to only two years, for both of those years GC prediction using correlated predictors (having intercorrelations  $> .45$ , see Appendix III) was numerically (though not significantly) better than GC prediction with uncorrelated predictors (when the sample size was the maximum; review Figures 1 and 2). It may be that given (1) the unique factor structure for intercorrelated predictors used and (2) the reliability of the intercorrelated predictors used, prediction using moderately intercorrelated predictors may be a desirable method.

#### Extreme Outcomes in Figure 5 and 6

As can be seen from Figures 5 and 6, respectively, two LP (local-proxy) calibration conditions produce substantially higher error means for 1984 and somewhat higher means for 1985 data:

BMGQ

(where      B= sample size= large  
               M= source of UGPA= MSU  
               G= range restriction= on  
               Q= predictor intercorrelation  $> .45$ )

and BNOQ

(where     B= sample size= large  
              N= source of UGPA= all  
              O= range restriction= off  
              Q= predictor intercorrelation >.45).

This underscores the previous recommendation that local-proxy calibration be performed with uncorrelated predictors due to the combination of (1) unreliability of the proxy criterion and (2) the moderate sample size. Beyond these observations, the question remains-- 'what affected prediction for these two years which was not evident for the preceding years?' Multicollinearity might be offered as an explanation for the unusual increase in error because, in both conditions, uncorrelated predictors are employed in the calibration. For the three years prior to 1984, however, the error level is uniformly low. Although the BMGQ calibration conditions were more restricted than were the data subsequently estimated by its selection formula, BNOQ (using unrestricted calibration conditions) exhibits greater error than BMGQ.

One explanation might be sampling error. Although the LP calibration would ordinarily include all cases in the calibration and in the estimation process, for this experimental study, calibration samples are, in fact, limited artificially to condition (a) of 80 cases, or to condition (b) of 135 cases, while the application sample is unrestricted. The 1984 and 1985 data entered the study midstream (it was added after 1981-1983 condition samples

had already been drawn), and because the case counts for these years were low, the restricted calibration samples for these years were somewhat under-sized (1984: B= 117, and 1985: B= 111, while for 1981-1983: B= 135). Although the calibration with the restricted sample may be expected to be less reliable than a calibration with a larger sample, the opposite outcome is also quite possible.

Although sampling error might account for this erratic prediction, such an explanation seems inconsistent with Figure 8, for instance, which depicts large and small sample prediction (BMOQ vs AMOQ) equal and constant across five years. Interactions only occur dramatically for 1984 data suggesting that a year effect is present for 1984 and perhaps (to a lesser extent) for 1985. The nature of this potential year effect is not evident. Because this apparent year effect is not systematically associated with any particular prediction factor, it can be suspected that it results from general unreliability in the selection equation due, perhaps, to some particularly incongruent non-MSU veterinary prerequisite UGPAs, or non-MSU, ordinary UGPAs which appear for these two years. Only these two variables are expected to be both important and yet potentially unreliable (due to possible differences in grading standards) to the degree sufficient to cause such a year effect. Probably the most feasible explanation for the resulting interactions are that some unusually incongruent non-MSU UGPAs in the 1984 (and to a lesser

extent, 1985) data when combined with two other prediction invalidating conditions (prediction factors) created notably unreliable regression equations. When the unreliable equations were used to create GGPA estimates, they again drew upon a UGPA predictor which remained quite unreliable. Therefore, in using a proxy criterion, it must be remembered that unreliability makes its mark both in the calibration of the equation and in the estimation of the criterion.

#### Cautions Regarding Study Realism

The small number (5) of suitable cohorts available for analysis limits the emergence and range of prediction factors in the data. To compensate for such limitation, this study exaggerated potential sources of variation in prediction error by deliberately biasing the selection formulas. This was achieved by (1) selecting on certain variables (eg. PVUGPA and MSU UGPA), (2) controlling the sample size, and (3) controlling the admission of correlated variables into the formula calibration. Unfortunately, some of the resulting conditions depart from realism. For that reason, omnibus tests of factor effects that test for general effects (over a large number of conditions) may have less practical validity than certain realistic, specific contrasts.

Some calibration conditions in this study are either not realistic to practice, or they might otherwise mislead

the interpreter of the study. In the method contrasts, for example, the sample sizes of the proxy methods are always greater than those of the GC (general-criterion) method, which accounts for the more erratic variation of GC error. Small sample size here was defined as data from only two of the three years of the cases available. Roughly 270 veterinary graduate cases were available to GC method calibration within three years of data, while about 900 cases were available to the GP (general-proxy) and to the GM (general-mixed) methods within the three years of data.

For the tests of methods, error was defined as error-in-ranking the estimated GGPA's against actual GGPA (or, LEWAR, where

L= natural log value  
 E= estimated  
 W= weighted  
 A= absolute value  
 R= error in rank).

Although the use of error in ranks was necessary in order to compare raw score regression error (for the proxy-criterion methods) with standardized regression error (for the GC method which required standardized regression in order to correct partial coefficients), rank error was more valid with regards to selection error. This weighted LEWAR and LEWA error

(where L= natural log value  
 E= estimated  
 W= weighted  
 A= absolute estimation error residual)

should be more appropriate than using ordinary unweighted

estimation error, as ordinary estimation error has little importance for cases above the cut-score. In addition, the use of log transformed error values gave greater weight to errors of large magnitude. Larger errors, were believed most likely to be due to (1) errors near the cut score, (2) errors from non-MSU UGPAs, or (3) criterion errors. Among generalizable calibrations, and among LP calibrations, (2) and (3) should remain constant, leaving cut-score region errors to explain the differences between these conditions. A problem linked to the use of weighted scores is the loss of degrees of freedom and the resulting liberalizing effect on significance tests; systematically weighting some cases greater than others, is the same (with respect to degrees of freedom) as deleting some cases.

In the LP (local-prediction) method contrasts, the calibration sample sizes (either 80 or 135) were less authentic due to limits in absolute number (it became impossible to provide even these numbers of cases for all conditions for the years 1984 and 1985). Error measures also were less realistic for the test across LP (local-proxy) methods, because they were in LEWA error (logs of weighted absolute differences between the estimated and actual criterion) rather than in ranking error. This sacrifice was made in the hope of improving the power of the statistical tests.

In practice, selection formulas would be applied to wide ranges of applicant scores; therefore, selection

formulas might ideally be calibrated on samples having the same wide ranges of scores. According to Table 3 (adapted from Richards, 1982), except for measurement error in the independent variable, raw-score regression coefficients could remain correct despite selection on the independent variable. Thus restriction of range might not be a problem if one wished to use a raw-score regression model as a selection formula. Such reliability in scores, however, is difficult to substantiate. Richards also demonstrates that (1) variation in the dependent variable range between application and calibration samples, and (2) variation in scale interval meaning affects the validity of the raw-score regression coefficient. The alternative regression procedure, standardized regression, is not affected by the problem of scale variation, and in addition, it allows for correction (actually, only a shrinking of error) of its betas (or partial correlation coefficients) for restriction of range on both the independent and dependent variables. Alexander et al. (1987) provide a formula for correcting correlations for both types of range restriction, and the author has extended the application to the betas (partial coefficients) in a standardized regression model. Since correction, in fact, is likely to be conservative (Linn, Harnisch, and Dunbar, 1981) and may ignore significant predictors (to result in an under-specified model), 'adjustment for dispersion' may be a more appropriate expression. In addition, such adjustment is also required



for expansion of range, where this occurs. In the practical application of selection formulas, the proxy-criterion methods would ordinarily not require any adjustment for dispersion differences between calibration and application samples, whereas, adjustment would be required for the conventional GC (general-criterion) method. Ordinarily, therefore, the error due to dispersion differences would be greater for the GC approach since the adjustment could only be approximate.

In this study, however, the validity-test sample happens to be ideal for the GC method because the application sample is range restricted (to upper distribution cases) in the same way as the GC calibration sample is restricted. The (unrestricted sample) proxy-criterion methods are faced with an uncharacteristic range restriction problem relative to the (range restricted) validity-test sample. In the test of methods, this has been partly countered by adjusting the GC (general-criterion) selection formulas for application to the total ranges of applicant scores. As such an adjustment is expected to be conservative, however, the GC method is expected to retain a slight predictive advantage attributable to the characteristics of the application sample of this particular study. On the other hand, the potential exists for the adjustment to accrue extraneous errors if the standardized regression is subject to the effects of sampling error and therefore imprecise

(standardized score regression is required in order to estimate the partial correlations to be adjusted for dispersion differences for each predictor). The author's adjustment technique, though extrapolated from respected bivariate correlation, has no literature to support its use with multiple coefficient correlation. This author did test the procedure on a small scale simulation with anticipated results, and therefore uses it with some justified confidence.

Of course, argument can be offered for not adjusting the GC selection formula, inasmuch as the full-range selection formula will not be optimal for cut-score situations where the selection ratio may be small. Such a course, nevertheless, risks the likelihood of misspecifying the selection model by erroneous predictor inclusions or exclusions or by weightings associated with the use of a restricted calibration sample.

#### Why $R^2$ Wasn't Used as a Measure of Validity

$R^2$  is the square of a multiple correlation coefficient, a conventional measure of validity for regression equations.  $R^2$  was not used in this study as an index of validity for several reasons. Not only does  $R^2$  not give greater weight to cases near the cut-score, but it gives greater weight to cases the farthest from the cut-score. Because  $R^2$  is a variance statistic, it follows that cases farther from the mean will contribute a disproportionate

share of the variance:

$$\text{variance} = \text{SUM} (\text{score} - \text{mean})^2 / N.$$

Note that the difference from the mean is squared, thus larger discrepancies from the mean can disproportionately influence the magnitude of the variance (e.g. the outlier problem). Because a graduate candidate distribution represents the upper tail of a distribution (higher ability college students), this distribution will be strongly negatively skewed with the distribution mean near the cut score. Under these circumstances, the most likely means of improving an  $R^2$  would be to improve prediction at the place in the distribution the most extreme from the mean: among the highest ability candidates. Such an 'improvement' in prediction may, in fact, decrease the level of discrimination in the cut-score region. It is also possible for the  $R^2$  to increase significantly without any corresponding change in the ordering of scores. In such circumstances, selection would remain unchanged despite better absolute prediction. Although the alternative index of validity, transformed prediction error, leads (as mentioned before) to liberal significance tests, it, nevertheless, can measure the change of interest to the admissions office.

### Regression Equations Have Superfluous Predictors

To control for confounding, calibration (regression) runs were executed without deleting ineffective predictors by means of conventional statistical tests. Statistical significance tests for coefficients are based partly on the number of predictors and partly on the size of the sample. Significance tests thus potentially confound with prediction factors such as (1) sample size and (2) intercorrelation of predictors. The consequence of superfluous predictors in the regression equation, however, is the addition of random error to the predicted scores, thereby attenuating actual predictor validities (Deegan, 1976). Equations without superfluous predictors will follow below.

### Practical Implications of the Study

Presented in Tables 22, 24, and 25 are regression models developed through a stepwise multiple regression procedure where conventional methods of coefficient significance testing ( $\alpha = .05$ ) have been used to refine the set of active predictor variables. In Table 22 (Method betas by condition) three estimation methods, GC (general-criterion), GP (general-proxy), and GM (general-mixed) use pooled data (1981 through 1983) to calibrate a single (estimation) regression equation for each method. For the LP method, 1984 and 1985 data are used to calibrate a regression equation for each of the two years. The GC and

GM methods (using GGPA and PVUGPA [veterinary prerequisites GPA] respectively) produce highly similar regression equations and they also give the greatest weight (.94, virtually all the predictive weight) to the UGPA variable. The GP method differs from the previous two methods in its weighting scheme, and weights UGPA at .77. The LP method runs (1984 and 1985) utilize fewer predictors and weight UGPA at .67 and .63 respectively.

The similarity between GC (general-criterion) and GM (general-mixed) calibrations further underscores the similarity of UGPA and GGPA at the site under study (although for GM the alternate criterion is, in fact a proxy, PVUGPA= veterinary prerequisites UGPA).

As displayed in Table 22, for the GC method, the selection formula validity (without crossvalidation) with five years of data was .91 (or about 80% of criterion variance accounted for by the predictors, or  $R^2 = .81$ ). The best average error using the GC formula to predict subsequent GGPA ranking of approximately 90 vet school graduates averages was a discrepancy of 17 positions. Also in Table 22, proxy-criterion methods obtained similar validity coefficients (non-crossvalidated). Of course these formulas predicted the proxy rather than the real criterion. Measured in terms of actual prediction error, the best average (absolute) proxy criterion (GM) prediction error was .29 from the actual GGPA.

Tables 24 and 25 provide additional LP runs for years 1981 through 1983; in Table 25 regression equations were calibrated on only 75 percent of available cases. Across the five years the weighting scheme varies somewhat, although UGPA maintains its dominance in prediction (ranging from .63 to .99 for the full samples).

Sample-size differences (100% vs 75%) have a minor effect on the regression coefficients estimated, but little effect on the predictors selected.

Because this study is site specific, generalization from these findings to other sites must be regarded as tenuous. One important feature of this data set is its homogeneity with regards to origin of UGPA (undergraduate GPA is predominately MSU). Admission of students with non-MSU UGPAs may require a relatively higher level of ability; graduate school faculty may be more willing to accept equivalent credentials from students whom they know rather than from less familiar candidates, thereby ruling out all but the highest performing non-local candidates. Should such be the case, prediction error in ranks would be minimized (rank error decreases at the extremes of the distribution). This could explain (1) why UGPA is so effective as a predictor, and (2) why non-MSU UGPAs are not associated with significantly greater prediction error (in ranks). At other sites or in other selection situations, the UGPA and GGPA may be less dependent, opening the potential for (1) alternative predictors with competitive

Table 22

Method betas under optimal prediction conditions

Predictor	<u>G C</u> $\beta$ s.e.	<u>G M</u> $\beta$ s.e.	<u>G P</u> $\beta$ s.e.	<u>LP84</u> $\beta$ s.e.	<u>LP85</u> $\beta$ s.e.
01 actach	-.032 .013	-.031 .013	-.043 .019	.	.
02 age	.	.	.	.	.093 .034
03 avgcred	-----	-----	-----	-----	-----
04 cred	.447 .119	.460 .120	.346 .164	.	.
05 ugpa	.938 .037	.940 .037	.767 .053	.674 .031	.626 .038
06 int1sc	.	.	-.149 .019	.	.
07 int2sc	.	-.027 .013	.	.	.
08 mcatb	-----	-----	-----	-----	-----
09 mcatc	.048 .018	.051 .018	.084 .022	.	.
10 mcatp	.038 .018	.057 .018	.	.	.
11 mcatq	-----	-----	-----	-----	-----
12 mcatr	-.039 .015	-.046 .015	.	.	.
13 mcats	-----	-----	-----	-----	-----
14 narr	-----	-----	-----	-----	-----
15 numterms	.122 .039	.128 .039	.	.	.
16 pfcred	.	.	.062 .019	.	.
17 pts	-.678 .114	-.692 .114	-.445 .164	-.110 .023	-.125 .034
18 pvspts	.217 .018	.213 .018	.170 .025	.341 .031	.383 .039
19 sex	-.049 .014	.	.	.	.
20 sumcred	-----	-----	-----	-----	-----
21 totcred	-----	-----	-----	-----	-----
22 vetexp	-----	-----	-----	-----	-----
23 workexp	-----	-----	-----	-----	-----
R	.91672	.91599	.81681	.93142	.91284

Note: See Table 23 for predictor variable definitions  
 GC= general-criterion, GP= general-proxy,  
 GM= general-mixed, LP84= local-proxy for 1984, LP85= local-  
 proxy for 1985

Table 23

## Predictor names and their definitions

NAME	DEFINITION
01 actach:	activities and achievements (non-academic)
02 age:	age in years
03 avgcred:	average term credits
04 cred:	total (grade point) credits
05 ugpa:	undergraduate grade point average
06 int1sc:	interviewer rating number 1
07 int2sc:	interviewer rating number 2
08 mcatb:	Medical College Admissions Test: Biology
09 mcata:	Medical College Admissions Test: Chemistry
10 mcatp:	Medical College Admissions Test: Physics
11 mcatq:	Medical College Admissions Test: Quantitative
12 mcatr:	Medical College Admissions Test: Reading
13 mcats:	Medical College Admissions Test: Science Reasoning
14 narr:	narrative writing sample
15 numterms:	number of terms enrolled in college
16 pfcres:	credits on a pass/fail basis
17 pts:	total honor points
18 pvspts:	total honor points in veterinary prereq.
19 sex:	gender
20 sumcred:	total veterinary prerequisites credits
21 totcred:	total grade-point and pass/fail credits
22 vetexp:	veterinary experience rating
23 workexp:	work experience rating



Table 24

LP betas by year: full sample

Predictor	<u>1981</u> $\beta$ s.e.	<u>1982</u> $\beta$ s.e.	<u>1983</u> $\beta$ s.e.	<u>1984</u> $\beta$ s.e.	<u>1985</u> $\beta$ s.e.
01 actach -----					
02 age	.138 .047	. .	. .	. .	.903 .034
03 avgcred -----					
04 cred	. .	.640 .190	.591 .186	. .	. .
05 ugpa	.823 .031	.994 .064	.914 .064	.674 .031	.626 .626
06 int1sc	-.068 .023	. .	. .	. .	. .
07 int2sc -----					
08 mcatb	. .	.076 .030	. .	. .	. .
09 mcatc	. .	.058 .028	.082 .025	. .	. .
10 mcatp	.103 .026	. .	. .	. .	. .
11 mcatq -----					
12 mcatr	. .	-.101 .027	. .	. .	. .
13 mcats -----					
14 narr -----					
15 numterms	.170 .078	. .	. .	. .	. .
16 pfcired -----					
17 pts	-.388 .068	-.797 .189	-.719 .192	-.110 .022	-.125 .034
18 pvspts	.220 .031	.196 .029	.261 .029	.341 .031	.383 .039
19 sex -----					
20 sumcred	. .	. .	.067 .023	. .	. .
21 totcred -----					
22 vetexp	. .	. .	. .	. .	-.056 .026
23 workexp -----					
R	.91595	.91893	.92455	.93142	.91284

Note: See Table 23 for predictor variable definitions

Table 25

LP betas by year: 75% sample

Predictor	<u>1981</u> $\beta$ s.e.	<u>1982</u> $\beta$ s.e.	<u>1983</u> $\beta$ s.e.	<u>1984</u> $\beta$ s.e.	<u>1985</u> $\beta$ s.e.
01 actach -----					
02 age	.	.	.	.	.036
03 avgcred -----					
04 cred	.	.	1.027	.	.
	.	.	.226	.	.
05 ugpa	.852	.814	1.003	.735	.662
	.035	.031	.072	.050	.040
06 int1sc	-.053	.	.	.	.
	.026	.	.	.	.
07 int2sc -----					
08 mcath	.	.069	.	.	.
	.	.033	.	.	.
09 mcath	.	.085	.078	.	.
	.	.031	.030	.	.
10 mcatp	.105	.	.	.	.
	.028	.	.	.	.
11 mcatq -----					
12 matr	.	-.107	.	.	.
	.	.030	.	.	.
13 mcats -----					
14 narr -----					
15 numterms	.332	.	.	.	.
	.080	.	.	.	.
16 pfcres -----					
17 pts	-.435	-.189	-1.171	-.350	-.150
	.080	.028	.235	.113	.037
18 pvspts	.201	.189	.281	.344	.364
	.034	.031	.034	.037	.040
19 sex -----					
20 sumcred	.	.	.102	.	.
	.	.	.027	.	.
21 totcred	.	.	.	.237	.
	.	.	.	.114	.
22 vetexp	.	.	.	.	-.064
	.	.	.	.	.029
23 workexp -----					
R	.91853	.92245	.91723	.92782	.92002

Note: See Table 23 for predictor variable definitions

validity, and (2) substantially better relative prediction for multiple regression prediction methods.

For selection situations similar to that in this study, the use of admission test scores is open to challenge. Admission test predictors obtain only marginally significant ( $\alpha = .05$ ) regression weights (see the MCATc, MCATp, and MCATr weights in the GC column, Table 24) when the UGPA predictor variable is already in the selection equation. Nevertheless, if the test score information is used by undergraduate institutions (as a secondary purpose for the data), to evaluate and modify internal curricular and grading standards, abandonment of admission test requirements might well result in a decline in UGPA validity. The potential influence of the admissions test on UGPA validity is probably sufficient reason to retain admissions test scores in the selection equation even though the immediate consequence may be marginal. Over the course of several years, the continuing inclusion of the admissions test predictors in the selection formula may preserve or even improve the validity of UGPA.

One strategy would be to use the admission test scores to correct all UGPAs, or to correct just the outside (e.g. non-MSU) UGPAs. The work of Linn (1966) in adjusting GPA was cited earlier. He found that GPA adjustment by admissions test scores effectively eliminated the GPA error due to source of GPA for high school GPAs. After correction of the appropriate UGPAs, admission-test

predictors would be withheld from the selection formula. In such a procedure, UGPAs would be adjusted up or down depending on the UGPA-admission test score discrepancy. Such a process, nevertheless, might provide little benefit to selection where few outside UGPAs reside near the cut score.

The inclusion of demographic variables in the regression equation may be quite informative for purposes of research. For application however, the use of demographic variable predictors cannot be recommended. This caution must be exercised because data samples are voluntary; hence, they are non-random data, plagued with systematic selection effects related to candidate recruitment and personal motivation. Effects attributed to categorical variables (e.g. race, gender) may be completely spurious. For example, due to a shortage of black candidates in human medicine, all minimally qualified blacks may be intensely recruited to human medicine, greatly depleting the remaining pool of blacks who would consider application to veterinary study. Blacks from this depleted pool would not provide a sample of black characteristics which could be validly generalized. Where data are not random, demographical variables must be regarded as moderator variables which may control variance to a certain extent, but which do not account for that variance in a literal sense (e.g. gender effect not being due to one's gender, per se). Ultimately, the predictive

advantage gained by blocking cases according to (moderator) categories (e.g. gender, race) may be retained by identifying other variables which account for the qualitative differences between category levels. For example, Table 22 shows a significant GC weight for the predictor variable "sex". In reality, the sex predictor variable may be moderating "level of affection for animals", or "temerity towards human patients". By including measures of these traits in the selection equation, the advantage of including the gender variable may disappear. Use of the categorical (moderator) predictor (e.g. gender) in the actual admissions selection formula merely on the basis of its predictive value is a questionable policy which is difficult to justify.

Beyond the consideration of merit, certain categorical variables neutral to race, religion, or gender (e.g. economic disadvantage) may be chosen and weighted by the admitting institution to create a non-merit criterion for admissions as an exception to the usual merit criteria (see Roos, 1978, for relevant information on non-merit admissions selection conforming to the Bakke judicial decision). The rating of such non-merit attributes, nevertheless, should not interfere with the evaluation of conventional applicant academic merit. The balance between merit and non-merit considerations should be specified as a consistent policy prior to application of non-merit considerations to particular cases.

## CHAPTER VII

### SUMMARY

Refinement of the selection process is an essential component of any serious effort to enhance public benefit from educational programs. In the introductory chapter it is noted that only recently has ease of data entry and retrieval made the prospects of a scientific graduate candidate selection process feasible. A review of research on candidate selection within the health sciences reveals a lack of consistent findings, likely partially due to limited sampling and sample sizes. Additionally, inconsistencies may arise from unspecified factors associated with particular years and locations which influence the composition of the non-random pool of applicants. One potential solution to such year and location effects would be to 'localize' estimation by attempting to estimate future performance based on a sample restricted to a single site and year. This prediction strategy was dubbed the LP (local-proxy) approach because it specified that the sample be local and that the regression criterion used be a proxy for the GGPA (graduate GPA). The proxy criterion would be the grade point average for the undergraduate veterinary prerequisite courses (PVUGPA), and this would allow the UGPA (undergraduate GPA) to serve as one of the multiple predictors. Three

additional variants of the LP approach would include

(1) the GP (general-proxy) where admissions data would be multi-year, (2) the GM (general-mixed) where admissions data would be multi-year (or multi-site) but the criterion would use either authentic or proxy criterions, or (3) the lone variable UGPA as an estimator of GGPA. Advantages to be gained by these methods might include (1) an increase in the sample size of local cases and (2) an expansion to the full applicant range of local cases (no cases need be deleted from the analysis for lack of a subsequent GGPA), (3) additional predictors can be added or deleted for any year (alternate admissions test scores could be accepted to a limited extent), and (4) previous GGPA data would not be required in order to estimate GGPA.

The research hypotheses called for some estimation methods to exceed the predictive validity of others:

- A: LP (local-proxy) to be more valid than UGPA
- B: GM (general-mixed) to be more valid than the GC (general-criterion)
- E: MSU applicant prediction to be more valid than prediction for non-MSU applicants,

or for the appearance of year or method effects:

- C: Year and method effects to appear in an analysis of methods
- D: Year effects to appear in an analysis of years and other factors

The review of literature in Chapter II concludes that research on predictor validities for the health sciences are not consistent from year to year nor across sites.

Linn's (1966) work to equate high school GPAs indicated that for these data the introduction of admissions test scores into the multiple regression was as effective as other more elaborate equating methods. Although some work has been done using UGPA as a proxy, no literature was found (other than the author's) where the context is academic selection. A large number of studies reported evidence of factors which moderate prediction among educational samples.

In the theoretical review of Chapter II topics of sampling, measurement, coefficient validity, multiple regression, and multivariate and univariate analysis of variance and the t-test are discussed.

In Chapter III conventional prediction is identified as general-criterion, or GC (general, because the regression equation is generalized across years and sites; criterion, because the criterion used is an authentic criterion). This prediction is compared and contrasted with the following three experimental multiple predictor prediction procedures:

(1) local-proxy, or LP (local, because the equation is not generalized beyond year or site; proxy, because the criterion is a 'stand-in' [veterinary prerequisites UGPA] for an 'authentic' criterion [graduate GPA]),

(2) general-proxy, or GP (general, because the equation can be generalized; proxy, because a proxy criterion is used),



(3) general-mixed, or GM (general, because the equation may be generalized; mixed, because the criterion will be an authentic criterion when available, otherwise, the criterion will be a proxy.)

Channels for the entry of error into the prediction process, potential for systematic and random error, advantages and disadvantages, and optimal conditions for the conventional and experimental approaches are discussed and compared.

Chapter IV describes the design of the proposed study. Five admissions cohorts were used from the Michigan State University College of Veterinary Medicine. Of these applicants, some were subsequently admitted to the veterinary program and subsequently received a graduate GPA. For all applicants, parallel admissions data were available. These took the form of grades, admissions test scores, ratings, and some demographic variables. Admissions data constituted the source of regression predictors, the graduate GPA was the authentic predictor, and the veterinary prerequisites GPA served as the proxy criterion.

Two major analyses were performed; both were repeated measures MANOVAs using dependent variables that were transformed prediction-error residuals. A test-of-methods MANOVA looked for methods and year effects, while controlling for the following factors: (Q) predictor intercorrelation, (A) sample size, and (N) source of UGPA.

The test-of-factors (and years) MANOVA used one estimation method, local-proxy, across five years of data. It looked for a year effect and it controlled for these factors: (Q) predictor intercorrelation, (A) sample size, (N) source of UGPA, and (O) restriction of range. For the test-of-methods MANOVA, general-criterion estimates were obtained with a standardized regression procedure, the betas being corrected for restrictions of range. For this MANOVA, prediction rank minus GGPA rank was the prediction-error used. For the test-of-years MANOVA, actual prediction-error from raw-score regression estimates was used.

Results are reported in Chapter V. As summarized in Table 12, Hypothesis D was confirmed. For Hypothesis D, cut-score prediction error differed across the five years. A view of Figures 5, 6, and 7, again reveals substantial interactions for 1984 data and modest interactions for 1985 data. Because dramatic interactions occur specific to only 1984, with lesser interactions for 1985, it is likely that data for these two years was of lower reliability (perhaps due to greater diversity in UGPA standards).

Unconfirmed are Hypotheses A, B, C, and E. For Hypothesis A, local-proxy estimation does not differ from estimation using UGPA as a sole predictor for cases in the cut score region. For Hypothesis B, both the general-criterion and the general-mixed method prediction validities are about the same. For Hypothesis C, neither method nor year effects occurs under the test-of-methods

for cases in the cut score region. Also failing to differ is prediction error for MSU and non-MSU applicants, contrary to Hypothesis E.

Chapter VI provides a discussion of the findings. While it is acknowledged that this study does not find a consistent predictive advantage for the proxy methods with the present data, some potential enhancements remain to be tried: (1) calibration of local-proxy equations using a single source of UGPAs (e.g. MSU), and (2) addition of random error to the UGPA predictor prior to the calibration. Also, potential for the methods may remain, albeit, under different circumstances. It is noted that the UGPA validity level was high for these data. In view of the geography of veterinary education which finds less than one school per state, this is not surprising. Where UGPA sources are more diverse, or where local grading standards are changing, the validity of the UGPA is bound to decline and competing estimation procedures (such as proxy criterion methods) may find practical use.

Intercorrelation of predictors did not appear to constitute a multicollinearity problem for general-criterion prediction. When the intercorrelated predictors did pose a problem for local-proxy estimation, it may have been due to the greater number of predictors tolerated in the high intercorrelation condition (and not from intercorrelation, per se). Kenny (1982) did note that multicollinearity was associated with predictor

unreliability, overlapping factors, and with the high correlation of the intercorrelated predictors with the criterion. Perhaps multicorrelation is not a serious problem for moderately intercorrelated admissions data which are sufficiently reliable and factor independent.

The reader is again cautioned regarding several aspects of the study which might be misleading: (1) general-criterion sample size levels were numerically smaller than those for the proxy-criterion methods; this was true to life but, nevertheless, sample size levels were not identical, (2) the dependent variable in the test-of-methods differs from that in the test-of-factors (the first is reported in rank-error, while the second is reported in actual error); the rank-error was more relevant to effect on cut-score cases, although the actual error allowed more powerful testing, (3) the general-criterion regression equations were computed using standardized regression and adjusted for restriction of range while the proxy-criterion estimation used (unadjusted) raw-score regression estimates, (4) the high predictive validity of UGPA for these data may not generalize to other sites or programs; UGPA validity is likely tied to the dominating proportion of MSU UGPAs in the applicant pool.

It is acknowledged that the regression equations used to estimate graduate grade point average in this study were not optimal, because they retained non-significant predictors. Non-optimal estimation allowed the testing of

sample size and predictor intercorrelation which would have been confounded with statistical testing, had it also occurred. To provide accurate regression equations for the methods in this study, therefore, regressions were run again with statistical testing of coefficients. The general-criterion and general-mixed equations were virtually identical, while all equations gave the dominate predictive role to the UGPA predictor.

## **APPENDICES**

**APPENDIX I**  
**Substitute Merit Values**

Table AI

Values substituted for missing merit values

VARIABLES	VALUE
activities and achievements (non-academic)	2.5
age in years	21.935
average term credits	14.612
total (grade point) credits	58.
interviewer rating number 1	20.483
interviewer rating number 2	20.711
Medical College Admissions Test: Biology	2.
Medical College Admissions Test: Chemistry	2.
Medical College Admissions Test: Physics	2.
Medical College Admissions Test: Quantitative	2.
Medical College Admissions Test: Reading	1.
Medical College Admissions Test: Science	2.
narrative writing sample	3.072
number of terms enrolled in college	10.38
credits on a pass/fail basis	7.2
total honor points	162.
gender	1.53
veterinary experience rating	4.833
work experience rating	2.26



APPENDIX II  
Error Weight by GPA Plot

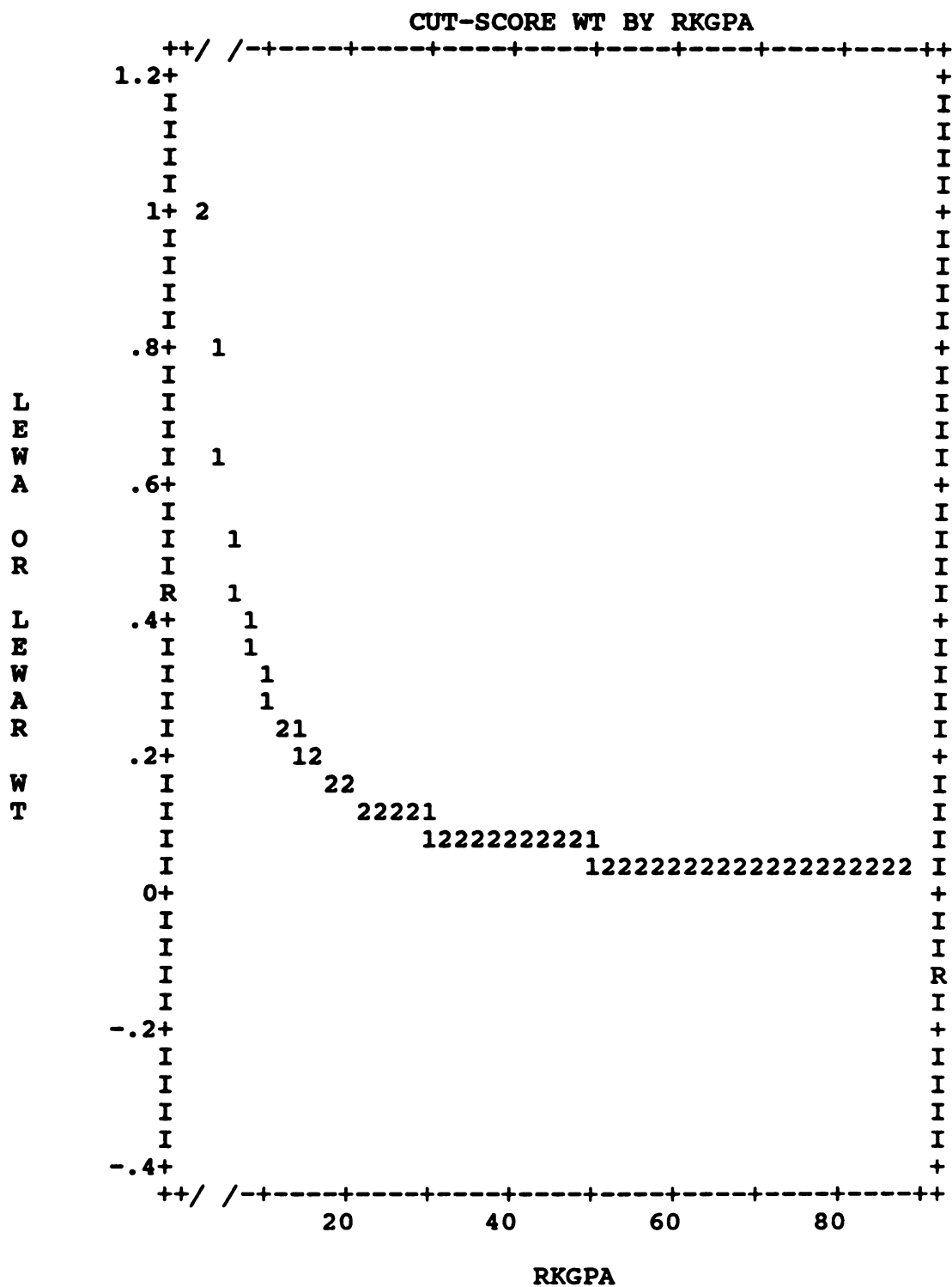


FIGURE IIA. LEWA (or LEWAR) weight plotted against RKGPA (n= 88).

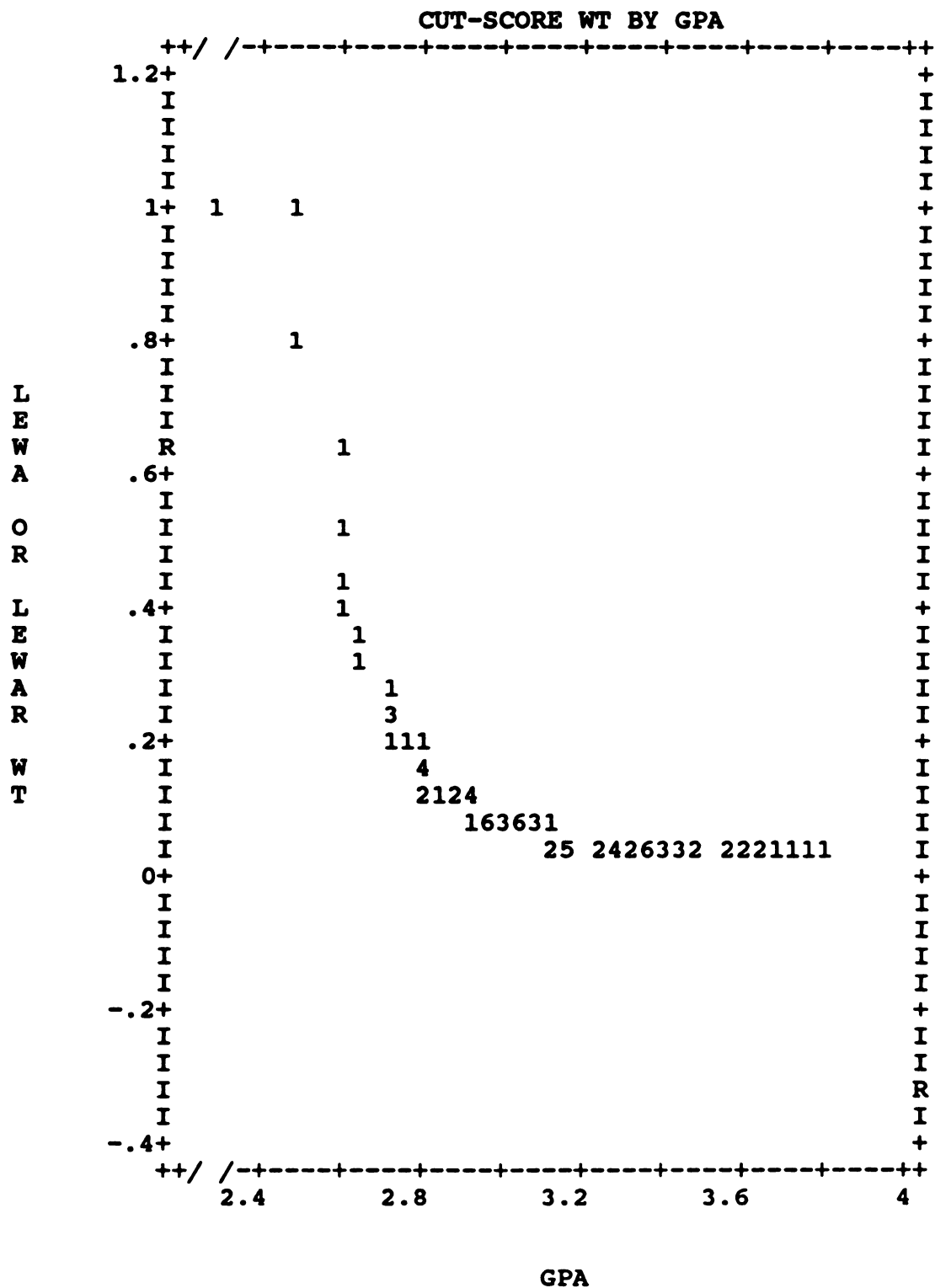


FIGURE IIB. LEWA (or LEWAR) weight plotted against GPA (n= 88).

### **APPENDIX III**

**Correlations: LEWAR-transformed Error by Predictors and GGPA**

Table AIII  
Correlations: LEWAR-transformed error by predictors and GGPA

ERROR		VARIABLE 2										
		N										
		PROB										
1984		ACTACH	AVGCRD	CUMGPA	MCATC	MCATR	MARR	PFCRED	PTS	PVSGPA	SEX	SUMCRED
LEWARMA		.0653 (.80) p=.283	-.1732 (.87) p=.054	-.1002 (.87) p=.178	-.2607 (.88) p=.007	-.1453 (.88) p=.088	-.1460 (.83) p=.094	-.0010 (.35) p=.498	.0229 (.87) p=.417	-.0791 (.87) p=.233	.0712 (.88) p=.285	.1185 (.84) p=.201
LEWARMB		.0386 (.80) p=.367	-.0512 (.87) p=.318	-.1495 (.87) p=.083	.0267 (.88) p=.403	.1381 (.88) p=.100	.0809 (.83) p=.282	-.1263 (.35) p=.235	.2107 (.87) p=.025	-.2243 (.87) p=.018	.0235 (.88) p=.414	-.0590 (.84) p=.336
LEWARMC		.1851 (.80) p=.050	.0013 (.87) p=.495	-.1731 (.87) p=.054	-.2644 (.88) p=.006	-.1043 (.88) p=.167	-.0782 (.83) p=.341	-.1534 (.31) p=.150	.0777 (.87) p=.237	-.2197 (.87) p=.020	-.1272 (.88) p=.118	.0048 (.84) p=.486
LEWARMD		.0770 (.80) p=.249	-.0782 (.87) p=.236	-.1532 (.87) p=.078	-.1880 (.88) p=.040	-.1933 (.88) p=.036	-.1053 (.83) p=.172	-.1314 (.35) p=.226	.0836 (.87) p=.221	-.1373 (.87) p=.102	-.0190 (.88) p=.430	.0141 (.84) p=.460
LEWARMI		.2105 (.80) p=.030	-.0430 (.87) p=.346	-.0937 (.87) p=.194	-.1631 (.88) p=.064	.0242 (.88) p=.411	-.1242 (.83) p=.132	-.0534 (.35) p=.380	.0890 (.87) p=.206	-.2203 (.87) p=.020	-.1262 (.88) p=.351	-.0531 (.84) p=.351
LEWARMJ		.2486 (.80) p=.013	-.0037 (.87) p=.487	-.1023 (.87) p=.173	-.1788 (.88) p=.048	.0443 (.88) p=.341	-.1382 (.83) p=.105	-.1016 (.35) p=.281	.1187 (.87) p=.137	-.2217 (.87) p=.020	-.1181 (.88) p=.141	-.1030 (.84) p=.226
LEWARMK		.1852 (.80) p=.050	-.0134 (.87) p=.451	-.0895 (.87) p=.261	-.2005 (.88) p=.031	.0176 (.88) p=.435	-.1589 (.83) p=.076	-.1975 (.31) p=.128	.1222 (.87) p=.130	-.2069 (.87) p=.027	-.0948 (.88) p=.180	-.1550 (.84) p=.132
LEWARML		.2033 (.80) p=.035	-.0217 (.87) p=.421	-.1109 (.87) p=.153	-.1931 (.88) p=.036	.0409 (.88) p=.353	-.1089 (.83) p=.164	-.3984 (.35) p=.009	.1607 (.87) p=.068	-.2449 (.87) p=.011	-.0622 (.88) p=.282	-.1934 (.84) p=.081
LEWARNA		.0813 (.80) p=.237	.1508 (.87) p=.082	-.0259 (.87) p=.406	-.0416 (.88) p=.350	.1752 (.88) p=.051	-.1634 (.83) p=.070	.1242 (.35) p=.239	-.0845 (.87) p=.218	-.1547 (.87) p=.076	-.0006 (.88) p=.498	.0498 (.84) p=.360
LEWARNS		-.0536 (.80) p=.319	.1055 (.87) p=.165	-.0074 (.87) p=.473	-.0937 (.88) p=.193	.1446 (.88) p=.086	-.1875 (.83) p=.045	.1917 (.35) p=.135	-.1418 (.87) p=.095	-.1446 (.87) p=.091	-.0317 (.88) p=.386	.0504 (.84) p=.356
LEWARNC		.1037 (.80) p=.180	.2071 (.87) p=.027	.0789 (.87) p=.235	-.0133 (.88) p=.451	.2621 (.88) p=.007	-.1179 (.83) p=.144	-.2595 (.35) p=.086	.1025 (.87) p=.172	-.0828 (.87) p=.196	.0184 (.88) p=.450	-.0005 (.84) p=.489
LEWARND		.1056 (.80) p=.176	.1594 (.87) p=.070	.0189 (.87) p=.431	-.0655 (.88) p=.272	.2343 (.88) p=.014	-.1225 (.83) p=.135	-.2490 (.35) p=.075	.1107 (.87) p=.154	-.1119 (.87) p=.151	-.0114 (.88) p=.456	.0112 (.84) p=.468
LEWARNI		.1701 (.80) p=.066	.0222 (.87) p=.170	.1037 (.87) p=.170	-.0931 (.88) p=.194	.2419 (.88) p=.012	-.1479 (.83) p=.081	-.2296 (.31) p=.092	.0077 (.87) p=.472	-.0329 (.87) p=.381	-.0183 (.88) p=.450	-.0242 (.84) p=.431
LEWARNJ		.1493 (.80) p=.083	.0158 (.87) p=.442	.0230 (.87) p=.416	-.0844 (.88) p=.217	.1453 (.88) p=.086	-.1809 (.83) p=.081	-.3103 (.31) p=.033	.0566 (.87) p=.301	-.1014 (.87) p=.176	-.1337 (.88) p=.108	-.0804 (.84) p=.262

Table AIII--continued

VARIABLE 2

ERROR

N  
PROB

	ACTACH	AVOCRED	CUMGPA	MCATC	MCATR	MARR	PFECRD	PTS	PVSGPA	SEX	SUMCRED
LEWARNP	-2160 (.93) P=.019	-1276 (.101) P=.102	-.0876 (.101) P=.251	-.0511 (.101) P=.306	-.0278 (.101) P=.391	-.1454 (.066) P=.079	.0181 (.38) P=.462	-.0402 (.101) P=.345	-.0056 (.101) P=.478	-.0289 (.101) P=.395	.1711 (.57) P=.102
LEWAROE	-1817 (.93) P=.041	-2143 (.101) P=.016	-.0485 (.101) P=.313	-.0689 (.101) P=.247	-.0341 (.101) P=.368	-.1547 (.066) P=.082	.1144 (.38) P=.247	-.0880 (.101) P=.191	.0454 (.101) P=.327	-.0452 (.101) P=.313	.1824 (.57) P=.087
LEWAROF	-2200 (.93) P=.017	-1742 (.101) P=.041	-.0400 (.101) P=.346	-.0543 (.101) P=.295	-.0227 (.101) P=.411	-.1243 (.066) P=.114	.0391 (.38) P=.408	-.0701 (.101) P=.243	.0504 (.101) P=.308	-.0087 (.101) P=.466	.1811 (.57) P=.086
LEWAROG	-1818 (.93) P=.041	-2157 (.101) P=.015	-.0862 (.101) P=.196	-.0580 (.101) P=.282	-.0477 (.101) P=.318	-.1257 (.066) P=.111	.0482 (.38) P=.332	-.1228 (.101) P=.111	.0162 (.101) P=.436	-.0722 (.101) P=.237	.1321 (.57) P=.164
LEWAROH	-2315 (.93) P=.016	-2080 (.101) P=.016	-.0478 (.101) P=.318	-.0463 (.101) P=.323	-.0235 (.101) P=.412	-.0812 (.066) P=.216	.0641 (.38) P=.351	-.0318 (.101) P=.378	.0419 (.101) P=.339	-.0081 (.101) P=.476	.1640 (.57) P=.111
LEWAROM	-2088 (.93) P=.022	-1344 (.101) P=.090	-.0908 (.101) P=.183	-.0136 (.101) P=.446	-.0070 (.101) P=.472	-.1606 (.066) P=.059	-.0075 (.38) P=.482	-.0135 (.101) P=.447	-.0178 (.101) P=.430	-.0058 (.101) P=.477	.1645 (.57) P=.111
LEWARON	-2375 (.93) P=.011	-1325 (.101) P=.083	-.0505 (.101) P=.308	-.0089 (.101) P=.461	.0555 (.101) P=.281	-.1027 (.066) P=.160	-.0064 (.38) P=.485	.0127 (.101) P=.450	.0119 (.101) P=.453	.0186 (.101) P=.422	.1593 (.57) P=.116
LEWAROO	-2531 (.93) P=.007	-1631 (.101) P=.053	-.0640 (.101) P=.262	-.0401 (.101) P=.345	.0082 (.101) P=.467	-.0952 (.066) P=.178	.0122 (.38) P=.271	-.0075 (.101) P=.470	.0246 (.101) P=.404	-.0557 (.101) P=.280	.1526 (.57) P=.129
LEWAROP	-2141 (.93) P=.020	-1504 (.101) P=.058	-.0796 (.101) P=.214	.0108 (.101) P=.457	.0637 (.101) P=.263	-.0897 (.066) P=.192	-.0335 (.38) P=.421	-.0080 (.101) P=.464	-.0133 (.101) P=.447	-.0115 (.101) P=.454	.1430 (.57) P=.144
LEWARPE	-1788 (.93) P=.043	-1683 (.101) P=.046	-.0634 (.101) P=.265	.0303 (.101) P=.382	.0582 (.101) P=.281	-.0647 (.066) P=.266	-.0570 (.38) P=.367	-.0367 (.101) P=.358	.0603 (.101) P=.375	-.0658 (.101) P=.254	.1325 (.57) P=.163
LEWARPF	-1624 (.93) P=.060	-2605 (.101) P=.004	-.0732 (.101) P=.233	.0000 (.101) P=.500	.0326 (.101) P=.373	-.0380 (.066) P=.353	-.0213 (.38) P=.456	.0286 (.101) P=.388	.0585 (.101) P=.281	-.0466 (.101) P=.322	.1229 (.57) P=.161
LEWARPG	-2223 (.93) P=.016	-1994 (.101) P=.023	-.0314 (.101) P=.378	.0455 (.101) P=.326	.0884 (.101) P=.190	-.0583 (.066) P=.286	.0512 (.38) P=.380	.0667 (.101) P=.254	.1011 (.101) P=.157	-.0058 (.101) P=.477	.1525 (.57) P=.126
LEWARPH	-1787 (.93) P=.043	-2167 (.101) P=.015	-.0453 (.101) P=.326	.0297 (.101) P=.384	.0688 (.101) P=.194	-.0326 (.066) P=.376	.0778 (.38) P=.260	.1041 (.101) P=.150	.0853 (.101) P=.198	.0148 (.101) P=.442	.1866 (.57) P=.108
LEWARPM	-2088 (.93) P=.022	-1344 (.101) P=.090	-.0908 (.101) P=.183	-.0136 (.101) P=.446	-.0070 (.101) P=.472	-.1606 (.066) P=.059	-.0075 (.38) P=.482	-.0135 (.101) P=.447	-.0178 (.101) P=.430	-.0058 (.101) P=.477	.1645 (.57) P=.111
LEWARPN	-2292 (.93) P=.014	-.0666 (.101) P=.254	-.0878 (.101) P=.165	-.0596 (.101) P=.277	-.0886 (.101) P=.189	-.1041 (.066) P=.156	.0047 (.38) P=.489	.0294 (.101) P=.385	.0031 (.101) P=.488	-.0378 (.101) P=.354	.1201 (.57) P=.187

Table AIII--continued

ERROR		VARIABLE 2		N		PROB							
ACTACH	AVGCRED	CUMGPA	MCATC	MCATR	NARR	PFCRED	PTS	PVSGPA	SEX	SUMCRED			
LEWARPO	-.2067 (.93) P=.023	-.0764 (.101) P=.224	-.0938 (.101) P=.175	.0317 (.101) P=.377	-.0887 (.101) P=.189	-.1214 (.96) P=.119	.0026 (.38) P=.494	.0372 (.101) P=.356	-.0095 (.101) P=.462	-.0420 (.101) P=.358	.1228 (.57) P=.181		
LEWARPP	-.2620 (.93) P=.006	-.0940 (.101) P=.175	-.0986 (.101) P=.163	.0639 (.101) P=.263	.1352 (.101) P=.088	-.0488 (.96) P=.319	.0382 (.38) P=.410	.0456 (.101) P=.325	.0085 (.101) P=.467	-.0108 (.101) P=.458	.1258 (.57) P=.176		
LEWARC	-.1688 (.93) P=.053	-.1554 (.101) P=.060	-.0251 (.101) P=.402	-.0631 (.101) P=.265	.0814 (.101) P=.209	-.1804 (.96) P=.039	.0511 (.38) P=.380	.0932 (.101) P=.177	.0434 (.101) P=.333	.0724 (.101) P=.236	.2183 (.57) P=.051		
LEWARP	-.1231 (.93) P=.120	-.0760 (.101) P=.225	-.2278 (.101) P=.011	-.1908 (.101) P=.028	-.0883 (.101) P=.190	-.1657 (.96) P=.053	.1484 (.38) P=.187	.0534 (.101) P=.298	-.0919 (.101) P=.180	-.0495 (.87) P=.1312	.0762 (.57) P=.267		
LEWARME	.1376 (.87) P=.102	-.1804 (.101) P=.035	LEWARNG	.0252 (.87) P=.408	GGPA	LEWAROM	.1382 (.87) P=.101	GGPA	LEWARPO	.1182 (.87) P=.138	GGPA	LEWARME	-.0448 (.101) P=.328
LEWARMF	.1477 (.87) P=.086	.0180 (.101) P=.428	LEWARNH	.0736 (.87) P=.249	GGPA	LEWARON	.0916 (.87) P=.199	GGPA	LEWARDP	.1330 (.87) P=.110	GGPA	LEWARMF	-.0423 (.101) P=.337
LEWARMG	.1275 (.87) P=.120	-.1798 (.101) P=.036	LEWARNM	.0859 (.87) P=.215	GGPA	LEWAROO	.1177 (.87) P=.139	GGPA	LEWARC	.0337 (.101) P=.368	GGPA	LEWARMG	.0006 (.101) P=.498
LEWARMH	.0556 (.87) P=.304	-.1671 (.101) P=.047	LEWARNN	.0875 (.87) P=.210	GGPA	LEWAROP	.0917 (.87) P=.199	GGPA	LEWARP	.1656 (.87) P=.063	GGPA	LEWARMH	-.1700 (.101) P=.045
LEWARMN	.0428 (.87) P=.347	-.1393 (.101) P=.082	LEWARNO	.1004 (.87) P=.177	GGPA	LEWARPE	.1698 (.87) P=.058	GGPA				LEWARMN	
LEWARMN	.1619 (.87) P=.067	-.0812 (.101) P=.210	LEWARNP	.1221 (.87) P=.130	GGPA	LEWARPF	.1511 (.87) P=.081	GGPA				LEWARMN	
LEWARMO	.1347 (.87) P=.107	-.1658 (.101) P=.049	LEWAROE	.1920 (.87) P=.037	GGPA	LEWARG	.1553 (.87) P=.074	GGPA				LEWARMO	
LEWARMPO	.1743 (.87) P=.053	-.1251 (.101) P=.106	LEWAROF	.1442 (.87) P=.061	GGPA	LEWARPH	.1171 (.87) P=.140	GGPA				LEWARMPO	
LEWARNE	.0664 (.87) P=.271	-.0433 (.101) P=.334	LEWAROG	.1858 (.87) P=.042	GGPA	LEWARPM	.1382 (.87) P=.101	GGPA				LEWARNE	
LEWARNF	.0376 (.87) P=.365	-.0728 (.101) P=.235	LEWAROH	.1340 (.87) P=.108	GGPA	LEWARPN	.1115 (.87) P=.152	GGPA				LEWARNF	

Table AIII--continued

VARIABLE 2												
N												
PROB												
1985	ACTACH	AVGCRD	CUMGPA	MCATC	MCATR	NARR	PFCRED	PTS	PVSGPA	SEX	SUMCRD	
LEWARME	.0420 (.93) P=.345	.2328 (.101) P=.010	-.1918 (.101) P=.027	-.1020 (.101) P=.155	.0717 (.101) P=.238	.1002 (.96) P=.166	-.1150 (.38) P=.246	.0136 (.101) P=.446	-.0968 (.101) P=.168	-.0022 (.101) P=.481	-.1104 (.101) P=.507	
LEWARMF	-.2809 (.93) P=.003	-.0439 (.101) P=.332	.0208 (.101) P=.416	.0519 (.101) P=.303	-.0085 (.101) P=.466	.0606 (.96) P=.276	-.1591 (.38) P=.170	-.0874 (.101) P=.192	.0329 (.101) P=.372	-.0076 (.101) P=.470	.0337 (.101) P=.402	
LEWARMG	-.0803 (.93) P=.222	.0354 (.101) P=.362	-.1838 (.101) P=.033	-.1500 (.101) P=.067	.0688 (.101) P=.247	-.1062 (.96) P=.152	.0311 (.38) P=.327	.0700 (.101) P=.243	-.1113 (.101) P=.134	.0636 (.101) P=.284	.1037 (.101) P=.577	
LEWARMH	.0300 (.93) P=.388	-.0389 (.101) P=.350	-.2037 (.101) P=.021	-.0469 (.101) P=.321	.0903 (.101) P=.185	-.1363 (.96) P=.093	.1337 (.38) P=.212	.1027 (.101) P=.153	-.1725 (.101) P=.042	.0287 (.101) P=.384	-.0038 (.101) P=.486	
LEWARMN	-.0873 (.93) P=.203	.1407 (.101) P=.080	-.1711 (.101) P=.044	-.1215 (.101) P=.113	.1701 (.101) P=.044	-.1106 (.96) P=.142	-.0875 (.38) P=.301	.0493 (.101) P=.312	-.1910 (.101) P=.028	-.0201 (.101) P=.421	-.0241 (.101) P=.420	
LEWARMN	-.1283 (.93) P=.110	.0447 (.101) P=.328	-.1055 (.101) P=.147	-.0015 (.101) P=.494	.1265 (.101) P=.104	-.1014 (.96) P=.163	.0689 (.38) P=.340	.0628 (.101) P=.266	-.1441 (.101) P=.075	.0048 (.101) P=.481	-.0050 (.101) P=.485	
LEWARMO	-.1477 (.93) P=.079	.0931 (.101) P=.177	-.1329 (.101) P=.093	-.0978 (.101) P=.165	.0812 (.101) P=.210	-.1331 (.96) P=.098	.1287 (.38) P=.221	.0485 (.101) P=.315	-.1605 (.101) P=.054	.0066 (.101) P=.474	.1486 (.101) P=.136	
LEWARMF	-.1606 (.93) P=.062	.0578 (.101) P=.283	-.0894 (.101) P=.187	-.1083 (.101) P=.140	.0788 (.101) P=.217	-.0765 (.96) P=.226	.1783 (.38) P=.142	.0195 (.101) P=.423	-.1131 (.101) P=.130	-.0187 (.101) P=.426	.1296 (.101) P=.168	
LEWARNE	-.2116 (.93) P=.021	-.1742 (.101) P=.041	-.0393 (.101) P=.348	-.1182 (.101) P=.119	.0706 (.101) P=.241	-.1066 (.96) P=.151	.1559 (.38) P=.175	.0283 (.101) P=.389	-.0212 (.101) P=.405	-.0031 (.101) P=.488	.1190 (.101) P=.186	
LEWARMF	-.2005 (.93) P=.027	.1220 (.101) P=.112	-.0515 (.101) P=.304	-.1295 (.101) P=.098	.0205 (.101) P=.420	-.1787 (.96) P=.041	.0422 (.38) P=.401	-.0106 (.101) P=.458	-.0488 (.101) P=.314	-.0401 (.101) P=.345	.1113 (.101) P=.205	
LEWARG	-.2356 (.93) P=.012	.1648 (.101) P=.050	-.1099 (.101) P=.137	-.1538 (.101) P=.062	.0468 (.101) P=.321	-.1303 (.96) P=.103	.1719 (.38) P=.151	.0098 (.101) P=.461	-.0482 (.101) P=.316	-.0413 (.101) P=.341	.1308 (.101) P=.166	
LEWARMH	-.2102 (.93) P=.022	.1223 (.101) P=.112	-.0726 (.101) P=.235	-.1284 (.101) P=.100	.0215 (.101) P=.416	-.1421 (.96) P=.084	.1690 (.38) P=.155	-.0394 (.101) P=.348	-.0438 (.101) P=.332	-.0367 (.101) P=.199	.1143 (.101) P=.199	
LEWARMN	-.1808 (.93) P=.041	.1578 (.101) P=.057	-.1198 (.101) P=.116	-.0881 (.101) P=.191	.0574 (.101) P=.284	-.1307 (.96) P=.102	-.0078 (.38) P=.481	-.0143 (.101) P=.443	-.0908 (.101) P=.183	.0298 (.101) P=.383	.1421 (.101) P=.146	
LEWARMN	-.2140 (.93) P=.020	.1087 (.101) P=.140	-.0945 (.101) P=.174	.0316 (.101) P=.377	.0365 (.101) P=.358	-.0917 (.96) P=.187	-.0912 (.38) P=.293	.0258 (.101) P=.396	-.0093 (.101) P=.463	-.0049 (.101) P=.481	.1091 (.101) P=.210	
LEWARNO	-.2027 (.93) P=.026	.1783 (.101) P=.037	-.1000 (.101) P=.160	-.0565 (.101) P=.287	.0456 (.101) P=.326	-.1282 (.96) P=.107	-.0300 (.38) P=.429	-.0136 (.101) P=.446	-.0578 (.101) P=.283	-.0771 (.101) P=.222	.1037 (.101) P=.577	



Table AIII--continued

ERROR

VARIABLE 2

N

PROB

	ACTACH	AVGCRD	CUMGPA	MCATC	MCATR	NARR	PFCRD	PTS	PVSGPA	SEX	SUMCRD
LEWARK	.1691 (.80) p=.067	-.0288 (.87) p=.396	.0887 (.87) p=.207	-.0697 (.88) p=.129	.1914 (.88) p=.037	-.1535 (.83) p=.083	-.1511 (.35) p=.193	.0391 (.87) p=.360	-.0593 (.87) p=.293	-.0516 (.54) p=.356	
LEWARNL	.1368 (.80) p=.113	-.0515 (.87) p=.318	.0544 (.87) p=.306	-.1355 (.88) p=.104	.1686 (.88) p=.058	-.1628 (.83) p=.071	-.1219 (.35) p=.243	.0682 (.87) p=.265	-.0942 (.87) p=.193	-.0177 (.88) p=.435	
LEWAROA	.0850 (.80) p=.201	.0785 (.87) p=.235	.0585 (.87) p=.285	-.0758 (.88) p=.241	.2196 (.88) p=.020	-.1140 (.83) p=.132	-.3614 (.06) p=.016	.0209 (.87) p=.424	-.1014 (.87) p=.175	-.0438 (.54) p=.342	
LEWAROB	.1072 (.80) p=.172	.1192 (.87) p=.136	.0449 (.87) p=.340	.107 (.88) p=.460	.2170 (.88) p=.021	-.0971 (.83) p=.181	-.2094 (.35) p=.114	.0098 (.87) p=.464	-.1137 (.87) p=.147	-.0982 (.88) p=.499	
LEWAROC	.1087 (.80) p=.169	.0741 (.87) p=.247	.0633 (.87) p=.280	-.0504 (.88) p=.321	.2121 (.88) p=.024	-.1078 (.83) p=.166	-.3918 (.35) p=.010	.0213 (.87) p=.422	-.0994 (.87) p=.180	-.0461 (.88) p=.335	
LEWAROD	.1084 (.80) p=.167	.0416 (.87) p=.351	.0445 (.87) p=.341	-.0399 (.88) p=.356	.2046 (.88) p=.028	-.1323 (.83) p=.117	-.1673 (.35) p=.168	.0185 (.87) p=.432	-.1275 (.87) p=.120	-.0551 (.88) p=.486	
LEWAROI	.1456 (.80) p=.089	.0275 (.87) p=.400	.0743 (.87) p=.247	-.0941 (.88) p=.192	.1634 (.88) p=.064	-.2007 (.83) p=.034	-.2672 (.35) p=.060	.0583 (.87) p=.296	-.0868 (.87) p=.212	-.0508 (.54) p=.213	
LEWAROJ	.1523 (.80) p=.089	.0244 (.87) p=.411	.0593 (.87) p=.293	-.0908 (.88) p=.200	.1236 (.88) p=.126	-.2262 (.83) p=.020	-.4320 (.35) p=.006	.0514 (.87) p=.318	-.0952 (.87) p=.190	-.0891 (.88) p=.282	
LEWAROK	.1466 (.80) p=.097	.0200 (.87) p=.427	.0690 (.87) p=.263	-.0860 (.88) p=.213	.1784 (.88) p=.048	-.1960 (.83) p=.038	-.4001 (.35) p=.009	.0115 (.87) p=.456	-.0816 (.87) p=.226	-.0564 (.54) p=.128	
LEWAROL	.1528 (.80) p=.088	.0014 (.87) p=.495	.0512 (.87) p=.316	-.0957 (.88) p=.187	.1433 (.88) p=.091	-.2195 (.83) p=.023	-.4623 (.35) p=.003	.0427 (.87) p=.347	-.1109 (.87) p=.153	-.0692 (.88) p=.261	
LEWARPA	.1410 (.80) p=.106	.0594 (.87) p=.302	.0513 (.87) p=.316	-.0432 (.88) p=.345	.1632 (.88) p=.064	-.1594 (.83) p=.075	.0236 (.35) p=.447	.1269 (.87) p=.121	-.1332 (.87) p=.106	-.0804 (.88) p=.321	
LEWARPB	.1301 (.80) p=.125	.1500 (.87) p=.083	.0644 (.87) p=.277	-.0420 (.88) p=.349	.2362 (.88) p=.013	-.1258 (.83) p=.129	-.0523 (.35) p=.383	.0855 (.87) p=.216	-.1450 (.87) p=.090	-.0629 (.54) p=.280	
LEWARPC	.1416 (.80) p=.105	.1554 (.87) p=.075	.0420 (.87) p=.350	-.0479 (.88) p=.329	.1863 (.88) p=.041	-.1599 (.83) p=.074	.0627 (.35) p=.360	.0491 (.87) p=.326	-.1271 (.87) p=.120	-.0476 (.54) p=.461	
LEWARPD	.1785 (.80) p=.057	.0938 (.87) p=.194	.0698 (.87) p=.268	-.0372 (.88) p=.365	.1856 (.88) p=.042	-.1871 (.83) p=.066	-.0361 (.35) p=.418	.0863 (.87) p=.213	-.1223 (.87) p=.130	-.0639 (.54) p=.277	
LEWARPI	.1351 (.80) p=.116	.1096 (.87) p=.156	.0038 (.87) p=.486	.0267 (.88) p=.402	.1812 (.88) p=.046	-.1926 (.83) p=.041	-.1757 (.35) p=.156	.0092 (.87) p=.406	-.1364 (.87) p=.104	-.1524 (.54) p=.078	

Table AIII--continued

VARIABLE 2 N PROB												
ERROR	ACTACH	AVGCRD	CUMGPA	WCATC	MCATR	NARR	PFCRED	PTS	PVSGPA	SEX	SUMCRD	
LEWARPJ	.1491 (.80) P=.093	.1108 (.87) P=.154	-.0299 (.87) P=.392	-.0727 (.88) P=.250	.1592 (.88) P=.069	-.2058 (.83) P=.031	-.2249 (.85) P=.087	.0219 (.87) P=.420	-.1741 (.87) P=.053	-.1232 (.88) P=.126	-.0594 (.84) P=.335	
LEWARPK	.1599 (.80) P=.078	.1793 (.87) P=.048	-.0033 (.87) P=.488	-.0335 (.87) P=.378	.1935 (.88) P=.035	-.2002 (.83) P=.035	-.2052 (.87) P=.118	-.0028 (.87) P=.490	-.1499 (.84) P=.083	-.1494 (.88) P=.082	-.0849 (.87) P=.371	
LEWARPL	.1811 (.80) P=.077	.1520 (.87) P=.080	-.0453 (.87) P=.339	-.0914 (.88) P=.199	.1550 (.87) P=.075	-.2192 (.83) P=.023	-.3206 (.85) P=.030	-.0079 (.87) P=.471	-.1848 (.87) P=.053	-.1207 (.88) P=.131	-.1267 (.84) P=.161	
LEWARC	.0251 (.80) P=.413	.0078 (.87) P=.471	-.0006 (.87) P=.337	-.0006 (.88) P=.498	.1068 (.88) P=.161	-.1658 (.83) P=.067	-.2114 (.85) P=.111	.0332 (.87) P=.380	-.0557 (.87) P=.304	-.1430 (.88) P=.092	-.0599 (.84) P=.333	
LEWARP	.0914 (.80) P=.210	-.0891 (.87) P=.206	-.0978 (.87) P=.184	-.2121 (.88) P=.024	.0439 (.88) P=.342	-.2200 (.83) P=.023	-.4541 (.85) P=.002	.1027 (.87) P=.172	-.1917 (.87) P=.038	.0773 (.88) P=.237	-.2014 (.84) P=.072	
LEWARNA	.0376 (.80) P=.380	-.3058 (.88) P=.002	LEWARNA (.88) P=.108	WORKEXP (.88) P=.007	GGPA (.88) P=.160	LEWAROA (.88) P=.046	WORKEXP (.88) P=.012	GGPA (.88) P=.106	LEWARPA (.88) P=.109	SEX (.88) P=.014	LEWARPA (.88) P=.050	
LEWARNB	-.1189 (.80) P=.171	-.3020 (.88) P=.002	LEWARNB (.88) P=.082	WORKEXP (.88) P=.146	GGPA (.88) P=.046	LEWAROB (.88) P=.046	WORKEXP (.88) P=.036	GGPA (.88) P=.086	LEWARPB (.88) P=.086	SEX (.88) P=.017	LEWARPB (.88) P=.106	
LEWARNC	-.1311 (.80) P=.143	-.3275 (.88) P=.001	LEWARNC (.88) P=.001	WORKEXP (.88) P=.007	GGPA (.88) P=.160	LEWAROC (.88) P=.049	WORKEXP (.88) P=.037	GGPA (.88) P=.081	LEWARPC (.88) P=.081	SEX (.88) P=.062	LEWARPC (.88) P=.092	
LEWARD	-.0851 (.80) P=.243	-.2645 (.88) P=.006	LEWARD (.88) P=.085	WORKEXP (.88) P=.085	GGPA (.88) P=.049	LEWARDOD (.88) P=.049	WORKEXP (.88) P=.037	GGPA (.88) P=.081	LEWARDPD (.88) P=.081	SEX (.88) P=.062	LEWARDPD (.88) P=.092	
LEWARI	-.1937 (.80) P=.035	-.3031 (.88) P=.002	LEWARI (.88) P=.006	WORKEXP (.88) P=.006	GGPA (.88) P=.079	LEWARIO (.88) P=.079	WORKEXP (.88) P=.025	GGPA (.88) P=.097	LEWARIPI (.88) P=.097	SEX (.88) P=.003	LEWARIPI (.88) P=.046	
LEWARNJ	-.1604 (.80) P=.086	-.2655 (.88) P=.006	LEWARNJ (.88) P=.086	WORKEXP (.88) P=.017	GGPA (.88) P=.054	LEWAROJ (.88) P=.054	WORKEXP (.88) P=.021	GGPA (.88) P=.082	LEWARPJ (.88) P=.082	SEX (.88) P=.008	LEWARPJ (.88) P=.022	
LEWARNK	-.1771 (.80) P=.074	-.2627 (.88) P=.007	LEWARNK (.88) P=.086	WORKEXP (.88) P=.006	GGPA (.88) P=.116	LEWAROK (.88) P=.116	WORKEXP (.88) P=.025	GGPA (.88) P=.087	LEWARPK (.88) P=.087	SEX (.88) P=.008	LEWARPK (.88) P=.051	
LEWARNL	-.1472 (.80) P=.115	-.2665 (.88) P=.006	LEWARNL (.88) P=.086	WORKEXP (.88) P=.028	GGPA (.88) P=.059	LEWAROL (.88) P=.059	WORKEXP (.88) P=.024	GGPA (.88) P=.072	LEWARPL (.88) P=.072	SEX (.88) P=.008	LEWARPL (.88) P=.021	

## REFERENCES

## REFERENCES

- Allen, M.J., Yen, W.M. (1979) Introduction to Measurement Theory. Monterey, California: Brooks/Cole.
- Alexander, R.A., Carson, K.P., Alliger, G.M., Carr, L. (1987) Correcting doubly truncated correlations: An improved approximation for correcting the bivariate normal correlation when truncation has occurred on both variables. Educational and Psychological Measurement, 47, p.309-315.
- Cattin, P. (1981) The predictive power of ridge regression: Some quasi-simulation results. Journal of Applied Psychology, 66:3, pp.282-290.
- Clapp, T.T., Reid, J.C. (1976) Institutional selectivity as a predictor of applicant selection and success in medical school. Journal of Medical Education, 51, pp.851-852.
- Deegan, J.Jr. (1976) The consequences of model misspecification in regression analysis. Multivariate Behavioral Research, April, 1976, pp.237-248.
- Doolittle, A.E., Cleary, T.A. (1987) Gender-based differential item performance in mathematics achievement items. Journal of Educational Measurement, 24:2, pp.157-166.
- Elliot, R., Strenta, A.C. (1988) Effects of improving the reliability of the GPA on prediction generally and on comparative predictions for gender and race particularly. Journal of Measurement in Education 25:4, pp.333-347.
- Goldman, R.D., Hewitt, B.N. (1976) Predicting the success of Black, Chicano, Oriental, and White college students. Journal of Educational Measurement, 13:2, pp.107-117.
- Goldman, R.D., Hewitt, B.N. (1975) Adaptation-level as an explanation for differential standards in college grading. Journal of Educational Measurement, 12:3, pp.149-161.
- Gough, H.G., Lanning, K. (1986) Predicting grades in college from the California Psychological Inventory. Educational and Psychological Measurement, 46, pp.205-213.
- Hakstian, A.R., Woolsey, L.K. (1985) Validity studies using the comprehensive ability battery (CAB): Predicting achievement at the university level. Educational and Psychological Measurement, 45, pp.329-341.

- Hart, M.E., Payne, D.A., Lewis, L.A. (1981) Prediction of basic science learning outcomes with cognitive style and traditional admissions criteria. Journal of Medical Education, 56:2, pp.137-139.
- Hogrebe, M.C., Ervin, L., Dwinell, P.L., Newman, I. (1983) The moderating effects of gender and race in predicting the academic performance of college developmental students. Educational and Psychological Measurement, 43, pp.523-530.
- Huberty, C.J., Mourad, S.A. (1980) Estimation in multiple correlation / prediction. Educational and Psychological Measurement, 40:1, pp.101-112, Spring.
- Humphreys, L.G., Taber, T. (1973) Postdiction study of the Graduate Record Examination and eight semesters of college grades. Journal of Educational Measurement, 10:3, pp.179-184.
- Huntsberger, D.V., Billingsley, P. (1973) Elements of Statistical Inference. Boston: Allyn and Bacon, pp.131-134.
- Jones, R.F., Thomas-Forgues, M. (1984) Validity of the MCAT in predicting performance on the first two years of medical school. Journal of Medical Education, 59:6, pp. 455-464.
- Keeves, J.P. (1988) Multivariate Analysis. In Keeves (ed) Educational research, Methodology, and Measurement: An International Handbook. Pergamon. New York, pp. 527-537.
- Kenny, D.A. (1979) Correlation and causality. Wiley & Sons, New York.
- Kirk, R.E. (1982) Experimental Design (2nd ed). Monterey, California: Brooks/Cole.
- Linn, R.L. (1966) Grade adjustments for prediction of academic performance: a review. Journal of Educational Measurement, 3, pp. 313-329.
- Linn, R.L. (1983) Pearson selection formulas: Implications for studies of predictive bias and estimates of educational effects in selected samples. Journal of Educational Measurement, 20:1, pp.1-15.
- Linn, R.L., Harnisch, D.L., Dunbar, S.B. (1981) Corrections for range restriction: An empirical investigation of conditions resulting in conservative corrections. Journal of Applied Psychology, 66:6, pp.655-663.

- Linn, R.L., Hastings, C.N. (1984) Group differentiated prediction. Applied Psychological Measurement, 8:2, pp.165-172.
- Loeb, J., Bowers, J. (1973) Programs of study as a basis for selection, placement, and guidance of college students. Journal of Educational Measurement, 10:2, pp.131-139.
- Markert (1983) Relationship of old and new MCAT scores to performance on the Part III examination of the NBME. Journal of Medical Education, 60:1, pp. 53-55.
- McCornack, R.L. (1983) Bias in the validity of predicted college grades in four ethnic minority groups. Educational and Psychological Measurement, 43, pp.517-522.
- McCornack, R.L., McLeod, M.M. (1988) Gender bias in the prediction of college course performance. Journal of Educational Measurement, 25:4, pp.321-331.
- Mehrens, W., Lehmann, I. (1984) Measurement and Evaluation in Education and Psychology (3rd Ed) Chicago: Holt, Rinehart, & Winston.
- Morris, J.D. (1986) Selecting a predictor weighting method by PRESS. Educational and Psychological Measurement, 46, p.853-869.
- Niedzwiedz, E.R., Friedman, B.A. (1976) A comparative analysis of the validity of pre-admissions information at four colleges of veterinary medicine. Journal of Veterinary Medical Education, 3:2, pp. 32-38.
- Neiner, A.G., Owens, W.A. (1985) Using biodata to predict job choice among college graduates. Journal of Applied Psychology, 70:1, pp.127-136.
- Neter, J., Wasserman, W., Kutner, M.H. (1985) Applied Linear Statistical Models. (2nd. Ed.) Homewood, Illinois: R.D.Irwin, Inc., p.10.
- Pedhazur, E.J. (1982) Multiple regression in behavioral research: Explanation and prediction (2nd. ed.). Chicago: Holt, Rinehart, and Winston.
- Richards, J.M. (1982) Standardized versus unstandardized regression weights. Applied Psychological Measurement, 6:2, pp.201-212.

- Roos, P.D. (1978) The implications of the Bakke decision on affirmative action admissions and related programs. In: Affirmative Action after Bakke. Connolly, W.B., Dilworth, E.J., and Leach, D.E. (chairmen). New York: Harcourt Brace Jovanovich, p.209.
- Ross, K.N. (1988) Sampling. In: Keeves (ed) Educational Research, Methodology, and Measurement: An International Handbook. Pergamon. New York, pp. 527-537.
- Sawyer, Richard (1986) Using demographic subgroup and dummy variable equations to predict college freshman grade average. Journal of Educational Measurement, 23:2, pp.131-145.
- Sawyer, R., Maxey, J. (1979) The validity of college grade prediction equations over time. Journal of Educational Measurement, 16:4, pp.279-284.
- Stuck, I.A. (1986) Selection by concurrent prediction: an alternative to the validity generalization of selection models. Michigan State University: Author.
- Tabachnick, B.G., Fidell, L.S. (1983) Using multivariate statistics. New York: Harper & Row.
- Thornell, J.G., McCoy (1985) The predictive validity of the Graduate Record Examinations for subgroups of students in different academic disciplines. Educational and Psychological Measurement, 45, pp.415-419.
- Wilson, K.M. (1982) A study of the validity of the restructured GRE aptitude test for predicting first-year performance in graduate study. Educational Testing Service Research Report 82-34, p.60.
- Wood, D.A., Langerin, M.J. (1972) Moderating the prediction of grades in Freshman engineering. Journal of Educational Measurement, 9:4, pp.311-320.
- Wright, R.J., Bean, A.G. (1974) The influence of socioeconomic status on the predictability of college performance. Journal of Educational Measurement, 11:4, pp.277-284.

MICHIGAN STATE UNIV. LIBRARIES



31293007776804