This is to certify that the

thesis entitled

A Multivariate Study of the Relationships Among
Types of Medical School Performance and Its Prediction

presented by

David B. West

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Educational Psychology

Lee S. Shulman
Major professor

Date August 14, 1979

MSU

OVERDUE FINES ARE 25¢ PER DAY
PER ITEM

Return to book drop to remove
this checkout from your record.

58 059
~~FEB 28 '81~~

~~MAY 2 '81~~
55 122

~~19 '81~~ 139

~~JUN 2 '81~~

23 170

69 ~~JUL 09 '81~~ R 190

~~JUL 23 '81~~

A MULTIVARIATE STUDY OF THE RELATIONSHIPS AMONG TYPES OF MEDICAL SCHOOL PERFORMANCE AND ITS PREDICTION

By

David B. West

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services,
and Educational Psychology

1979

ABSTRACT

A MULTIVARIATE STUDY OF THE RELATIONSHIPS AMONG TYPES OF MEDICAL SCHOOL PERFORMANCE AND ITS PREDICTION

By

David B. West

The principal research questions investigated by the study concerned the relationships among student performance in courses in osteopathic medical school. Using previous research and theory, four performance structures were identified: (a) a single, general factor structure; (b) a two-factor structure consisting of clinical and basic science performance factors; (c) a three-factor structure based upon course content and sequence; and (d) a four-factor structure which was formulated primarily on the basis of course content. Summary measures of student course performance were collected for students in two entering classes in a college of osteopathic medicine.

The relationships among the facets of performance specified by the models were estimated by using maximum likelihood confirmatory factor analysis. This technique permits the researcher to specify on which latent factors the observed performance scores should load, and provides maximum likelihood estimates of the factor loadings, the

correlations among true scores on the factors, and the unique variances of the observed variables. The four factor model, which included separate basic science and clinical skills performance factors, fit the observed relationships among the course performance measures best. The estimated correlations among true scores on the four factors were all high (i.e., .77 to .93) and highly statistically significant. Surprisingly high was the estimated correlation between true scores on the basic science and clinical skills factors (.927). Using the data from the second entering class, these relationships were successfully cross-validated.

Three "clusters" of independent variables were chosen to be used to predict performance on these four criterion performance factors. These clusters were: measures of science aptitude and achievement (i.e., Science GPA, Science MCAT score, and Biology GPA); verbal ability and achievement (i.e., MCAT Verbal score and English GPA); and behavioral science achievement (Behavioral Science GPA). The results of a multivariate regression analysis showed that performance on the criterion measures was significantly related to the science predictors. The results of the univariate regression analyses showed that the science predictors accounted for 63% of the variance on the Basic Science performance factor and over 40% of the variance on each of the clinical performance factors. While the associations between the same predictor and criterion variables were also highly significant in the cross-validation sample, the magnitudes of the associations were smaller.

The principal conclusions of the study can be summarized as follows:

1. Medical school performance is consistent across both subject matter domains and methods of measurement.

2. Consistent with the results of recently reported studies, basic science and clinical performance are more strongly related than had been reported in earlier studies.

3. Given a population with a wide variation in scores on objective predictors of medical school performance, these measures are substantially related to subsequent medical school performance in the pre-clinical clerkship years.

To all my friends and colleagues

who supported and encouraged

me in this effort.

ACKNOWLEDGMENTS

To the members of the committee, Lee Shulman, Bill Schmidt, Fred Tinning, John Schweitzer, Ed Smith, and Joe Byers, for the unique part each has played in my educational, personal, and professional development.

To Assistant Dean Fred Tinning, Dean Myron Magen, Dr. Ron Markert, Dr. Frank Bernier, and the faculty of the College of Osteopathic Medicine for their unexcelled cooperation and support.

To my mother and father, Betty and Joe West, who nurtured an enduring passion for learning.

Finally to Jim Mullin, Sandra Meacham, Jay and Jim Stratoudakis, Steve Downing, Fred Tinning, Sherri Sutton, and Bob Yinger for their encouragement, companionship, and support during the last six years.

* * * * *

TABLE OF CONTENTS

LIST OF TABLES

Table		Page

Table		Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

A medical degree offers entry into a socially relevant, high status, and potentially lucrative career. Since the attrition rate among medical students is low compared with other graduate and professional fields (i.e., about 5 to 7% nationally during the last 10 years), admission to medical school virtually guarantees entry into the allopathic (M.D.) or osteopathic (D.O.) medical professions. The problem faced by medical school faculty members and medical school administrative officials is this: Out of the approximately 2,000 to 4,000 applicants to the average medical school, who should be admitted to medical school and, in a probability sense, be virtually guaranteed entry into the medical profession.

This problem has been viewed through a number of frames of reference. The public policy view of the problem has been manifested in legislative hearings (led by, among others, Senator Edward Kennedy) which have been concerned with the shortage of primary care physicians and with the specialty and geographic distributions of physician manpower. (An entire federal bureau, the Bureau of Health

Manpower, has also been set up to deal with these concerns.) These national and state policy concerns are considered to be justified as the major portion of the costs of the student's medical education is paid by the state and federal governments and not by the student (Institute of Medicine, 1973).

The social aspects of the problem of the distribution of educational and career opportunities have been seriously argued in the recent Bakke case before the U.S. Supreme Court, in the editorial commentary on the case, and behind closed doors in medical school faculty and administrative conclaves. Finally, Dan Rather of CBS TV's "Sixty Minutes" irreverently offered an arresting economic perspective on the topic when he satirically noted that a medical degree is equivalent to a license to print money.

A significant input into the solution of the problem of whom to admit is the character of the medical school curriculum. Following the European model of the time, the first American medical schools' curricula consisted of two years of basic science education and two years of apprenticeship training in clinical medicine. A pursual of the current number of the Association of American Medical Colleges Curriculum Guide (AAMC, 1977) reveals that this is still the predominant curriculum model today.

The European model received considerable support from a study of American medical colleges by Abraham Flexner (1910) at the turn of the century. Flexner and his staff visited scores of public and private medical schools and investigated their facilities, faculty, and instructional practices. The Flexner Report whose influence is still being felt today concluded that the most educationally sound colleges were those whose curricula followed the scientifically based European model. Almost as a consequence, the medical schools which survived the onslaught of revelations of inadequate facilities and teaching practices made public by the Report were those schools which had strong programs in basic science education followed by adequately staffed clinical apprenticeships in well equipped hospitals with sufficient numbers of patients.

While the ultimate goal of the admissions process is to select applicants who will become good physicians, there remains a great deal of controversy within the medical profession about what a "good" physician is and how to assess this. Consequently, congruent with Nobel Prize winner Herbert Simon's theory of administrative decision making (Simon, 1957), admissions committee members have historically narrowed their collective problem space to a more manageable task: Based upon the simplifying assumptions that the medical school curriculum is fixed and that the first step toward an M.D. or D.O. degree is to complete medical school, they have essentially narrowed their concerns to selecting applicants who, in the view of the committee members, demonstrate a high probability of completing medical school. Committee members thus place a great deal of weight on objective measures of the applicant's general ability, and his or her aptitude and achievement in the sciences. These admissions variables have traditionally included premedical grade point average (GPA), premedical science GPA, and Medical College Admissions Test (MCAT) scores (Gough, 1971).

This pragmatic solution to the problem of who to admit has produced several years of research on which variables best predict academic success in medical school. The findings of this body of research will be discussed in more detail in Chapter II. However, the general findings can be briefly summarized as follows: Cognitive selection variables (i.e., GPA's and MCAT scores) are moderately related to classroom performance, show little relationship with clinical performance within medical school, and virtually no relationship with post-graduate clinical performance. However, since no satisfactory predictors of clinical performance have yet been found, the problem of who to admit is still generally solved by admitting applicants with relatively high GPA's and MCAT scores.

Committee members who oppose this traditionally strong emphasis on MCAT scores and GPA's assert that such an emphasis has resulted in the selection of overly scientifically-oriented applicants who later enter specialty and research careers rather than careers in primary care medicine (e.g., Gough, 1979). They also argue that this emphasis has caused the rejection of humanistically-oriented applicants, who in spite of their less impressive credentials in science, could be educated to become competent, patient-oriented, general practitioners (Gough, 1971; Rhoads, Gallemore, Bianturco, & Osterhout, 1974). This group, therefore, advocates placing more weight upon noncognitive measures of applicants' suitability such as motivation, problem solving ability, commitment to society. Advocates of the admission of additional numbers of minority applicants have similarly argued that minority applicants who could have become good general practitioners have also been denied entry into medical schools because of lower GPA's and MCAT scores than their middle-class, Anglo-Saxon counterparts.

The Problem

All groups agree that education and training in the basic sciences, clinical medicine, and clinical skills are necessary facets of medical education. However, applicants are selected primarily on the basis of objective measures which are related to performance in the first two years of the curriculum only. The crux of the problem, therefore, lies in the strength of the relationship between performance in the basic sciences and performance during the clinical portion of the curriculum. If no strong relationship between these two facets of medical school performance exists, the current practice of selecting applicants on the basis of measures of scientific aptitude and achievement should be re-evaluated as suggested by the "non-traditionalist" members of the admissions committee. On the other hand, if such a relationship does exist, two possible explanations of it warrant further investigation. First, as suggested by the European-Flexner curriculum model, adequate knowledge of basic science is a prerequisite to the acquisition of clinical principles and skills. Second, students who are able, or who possess a better base of knowledge in one of these areas can be predicted to perform better in the other area(s). In factor analytic terms, different types of performance can be conceptualized as either loading on a single factor or loading on a number of substantially correlated (i.e., oblique) factors.

As will be discussed more fully in Chapter II, reasonably strong relationships between basic science performance and clinical performance during medical school have been reported in the literature. Gough, Hall, and

Harris (1964) reported moderate correlations among yearly GPA's for a large sample of University of California medical students. Sirotkin and Whitten (1978) found sizeable canonical correlations between measures of performance in the following three areas: (a) basic science, (b) clinical medicine associated with different organ systems in the body, and (c) ratings of clinical clerkship performance. Finally, Maatsch, Downing, Sprafka, and Holmes (1978) factor analyzed scores on objective tests of clinical and basic science knowledge, and ratings of performance in clinical simulations. They found that all of their measures loaded on a single factor which accounted for approximately 40% of the variance of the correlation matrix of scores on the measures.

Purpose

Using the above information, three models of the structure of medical school performance can be hypothesized:

1. A two factor structure consisting of uncorrelated basic science and clinical performance factors.

2. A two or more factor structure in which the factors are correlated (e.g., Sirotkin & Whitten, 1978).

3. A one factor structure in which both clinical and basic science performance measures load on a single, general factor (e.g., Maatsch et al., 1978).

The first purpose of the current study is to evaluate the adequacy of these hypothesized performance structures to model the observed covariation among measures of medical school performance in basic science, clinical medicine, and clinical skills courses. Second, once an appropriate model has been selected, the strength of the relationships among the performance factors identified by the model can be estimated. The third purpose of the study is to estimate the relationship between the medical school performance factors, on the one hand, and typically used medical student selection variables (such as MCAT scores and premedical GPA's), on the other.

Method of the Study

When a researcher knows *a priori* what structures he or she wishes to investigate an ideal technique for doing so is covariance structure analysis (Jöreskog, 1974; Wiley, Schmidt, & Bramble, 1973). A related technique which is appropriate for the analysis of relationships among continuously scaled measures is confirmatory factor analysis (Jöreskog & Lawley, 1968). Confirmatory factor analysis allows the data analyst to specify on which factors (e.g., clinical or basic science) the performance measures load, and then to test the fit of this hypothesized structure to a set of data. The COFAMM program for confirmatory factor analysis developed by Sörbom and Jöreskog (1976) provides maximum likelihood estimates of the factor loadings, the correlations among the factors, and the unique variances of the measures. These estimates of the correlations among the factors can be used to test the hypothesized relationships between clinical and basic science performance discussed above.

In order to investigate these hypothesized performance structures, course grades and measures of clinical performance were collected for two entering classes of students matriculating at the Michigan State University College of Osteopathic Medicine. Also collected were students' MCAT scores, premedical GPA's, interview ratings, and other variables used to select applicants for admission to the College. These performance measures and admissions variables are described in Chapter III.

Using confirmatory factor analysis, the statistical models underlying the hypothesized performance structures can be estimated and the fit of each model to the covariation among the measures of student performance can be tested. The fit of the models can then be compared by using sequential chi-square tests for goodness of fit (e.g., Goodman, 1972). Once the most theoretically and statistically appropriate structure has been chosen, the multivariate relationship between the performance factors and typically used medical school selection variables can be estimated by using multivariate multiple regression (e.g., Finn, 1974).

The research approach outlined here offers at least two important advantages over those used in previous studies. First, while previous research studies have been concerned with one or the other, this study will investigate both the interrelationships among facets of performance during medical school _and_ the relationship between these performance factors and typically used selection variables. Second, in contrast to most studies in this area which have paradoxically employed bivariate correlational techniques to estimate the relationships among several variables (see Chapter II), this study will employ more statistically appropriate and powerful multivariate analytic tools to study these relationships.

Overview of the Dissertation

In Chapter II the little research which has investigated the interrelationships among different facets of medical school performance will be reviewed. Selected studies of the prediction of medical school classroom and clinical performance, and studies of performance hierarchies which may prove relevant to the current research problem will also be reviewed and discussed. Chapter III contains a description of the performance and admissions measures which were collected and a discussion of the data analytic methods to be used in analyzing the results of the study.

In Chapter IV the results of the estimations and tests of the hypothesized performance structures, and their relationships with the admissions variables will be presented and discussed. Finally, in Chapter V the relationships between the results of this investigation and the results of previous studies will be compared and the implications of this entire body of results for the distribution of educational and career opportunities for all groups in society will be discussed.

CHAPTER II

REVIEW OF RELATED RESEARCH

In this chapter, relevant research literature concerning investigations of the structure of medical school performance and its prediction will be discussed. All of the studies of which the author is aware which have investigated relationships among various types of performance in medical school have been reviewed in that section. Rather than presenting an annotated bibliography of research on prediction of medical school performance (about 200 such studies have been published) and investigations of performance structures in non-medical areas (which have been reviewed in books on factor analysis), an effort has been made to selectively review research in detail which has special substantive and/or methodological significance to the research problems under investigation in this study. The general findings of other studies will be summarized.

Investigations of the Structure of Academic Performance

General Findings

Pioneers in the field of mental measurement developed the methodology of factor analysis in order to account for the relationships among performance measures in terms of a few underlying or latent dimensions. The success of the technique is determined by how well the system of latent variables or factors reproduces the observed correlations among the performance measures being analyzed. Expanding upon the correlational techniques originated by Galton and Pearson, Spearman (1904) factor analyzed a set of ability measures and reported that the correlations among these measures could be accounted for by a single factor which he called "*g*" (for "general ability") and specific factors containing variance which was unique to each of the measures (so-called "unique factors"). Spearman (1927) defined *g* as the ability to see relationships. According to Jensen (1969), this ability is the quintescence of many traditional and contemporary definitions of general intelligence.

One traditionally used test of how well the correlation matrix has been reproduced by the factor or factors which have been extracted is to look at the residual covariance matrix among the measures after the hypothesized number of factors has been extracted. If these residual covariances are non-significant, it is generally concluded that no more factors can be extracted. This implies that no additional dimensions are needed to explain the observed correlations or covariances among the measures.

Later factor analyses of scores from ability tests which were similar to those originally analyzed by Spearman yielded significant residual covariances among the variables after the variance due to *g* had been statistically controlled. When additional factors were extracted, factors on which clusters of related abilities loaded were yielded by the analysis. British factor analysts have called these clusters "group factors." From their analyses of the grades of elementary and secondary school children, British factor analysts have consistently reported the existence of verbal (e.g., Composition, Literature, History, Geography), numerical (e.g., Arithmetic, Geometry), and practical (e.g., Handiwork, Drawing, Penmanship) group factors in addition to a general factor (Vernon, 1961).

Spearman's two-factor theory and group-factor theory are schematically compared in Figure 2.1. In group-factor analysis each variable loads on the general factor and on one (or occasionally more than one) group factor. Typically the general factor accounts for a plurality of the variance of the correlation matrix and the group factors for decreasing amounts of variance. For example, Burt

1. Two factor analysis			2. Group factor analysis						3. Multiple factor analysis					
Test	General factor	Specific factors	Test	General factor	Group factors A	B	C	Specific factors	Test	Multiple factors A	B	C	D	Specific factors
1	+	+	1	+	+			+	1	+				+
2	+	+	2	+	+			+	2	+				+
3	+	+	3	+	+			+	3	+	+			+
4	+	+	4	+	+			+	4	+				+
5	+	+	5	+		+	+	+	5		+		+	+
6	+	+	6	+		+		+	6	+	+			+
7	+	+	7	+		+		+	7		+			+
8	+	+	8	+			+	+	8			+	+	+
9	+	+	9	+			+	+	9			+		+
10	+	+	10	+			+	+	10			+		+

Figure 2.1 Comparison of factor analytic models (adapted from Vernon, 1961, Figure 1, p. 18).

(1939) performed a group factor analysis on the grades of elementary school children. The general factor accounted for 28% of the total variance in the correlation matrix and the verbal, numerical, and practical group factors together accounted for an additional 21% of the variance.

In contrast to the British approaches which involved extraction of a general factor first, American factorists led by Thurstone derived methods which yielded a number of what Thurstone called multiple factors. Thurstone's centroid technique involved successively extracting a number of factors and then rotating the reference axies ("factors") so that particular variables would load maximally on one factor and have negligible loadings on other factors. Using the centroid technique and subsequent rotations of the reference axies to what he called "simple structure," Thurstone (1938) factor analyzed 56 ability tests given to 240 college students. The result as with factor analyses of personality variables was a series of multiple factors (Vernon, 1961). Thurstone called these rotated factors "primary mental abilities" and strongly argued against Spearman's conception of a single general ability factor.

Conclusions about the structure of academic performance depend upon which factor analytic technique has been used. When a general factor has been extracted and no further analysis of the residual matrix is performed, Spearman's two-factor solution will result. When additional factors are extracted from the residual matrix, a group factor structure results. When a number of factors are extracted and rotated, a multiple factor solution emerges.

The intellectual and statistical challenge offered by factor analysis is that the same set of data can be analyzed by using any of the methods discussed above, and while conclusions based upon the results will differ, all of the solutions are mathematically "legitimate." An additional complexity results when the factors are rotated. Statistically, what is being done when rotations are performed is that the variance due to the first or general factors of a set of achievement measures is being redistributed among the group or multiple factors which have been created through rotation (Vernon, 1961). When an oblique rotation is performed (i.e., the factors are permitted to correlate), the factors are often at least moderately correlated implying an underlying general factor (Wolfle, 1940).

Any set of factors can be orthogonally rotated in an infinite number of ways, thus producing a theoretically infinite number of mathematically legitimate patterns of factor loadings. Statisticians (e.g., Lawley & Maxwell, 1971) refer to this as the problem of "indeterminancy."

One logical way of dealing with this problem of the lack of a uniquely identified solution is to specify, in advance on which factors the variables should load and then proceed to test this hypothesized performance structure (Jöreskog & Lawley, 1968). An early technique used for this purpose was Burt's multiple-group factor analysis (Harman, 1968; Hunter & Gerbing, Note 1). A later, more flexible technique is Jöreskog's confirmatory factor analysis. In both of these techniques the factors on which the variables load are specified in advance eliminating the rotation phase of the analysis.

Most of the studies of the structure of achievement have been done on intellectually heterogeneous groups at pre-university levels of education (e.g., elementary school children, military service personnel) and are, thus, not as generalizable to the research problems in this study as would be desirable. Studies which exemplify the research done on college and professional student populations will now be discussed.

Schoenfeldt and Brush (1975) calculated the GPA's in 12 subject matter areas (e.g., Humanities, Biological Science, Social Science) for over 1,900 undergraduate students. These 12 GPA's were then factor analyzed along with the student's high school GPA and Scholastic Aptitude Test (SAT) verbal and math scores. After a varimax rotation was performed, the analysis yielded three factors: (a) a general academic achievement factor on which 10 of the 12 GPA's loaded; (b) a factor consisting of grades in applied areas (i.e., Agriculture and Education); and (c) an SAT factor. From these results, the researchers concluded that college achievement is essentially a unitary trait.

In a similar study of law school grades, Boldt (1973) factor analyzed law school grades for 116 students and tested the goodness of fit of one, two, three, and four factor solutions. He similarly concluded that the matrix of law school grades consisted of essentially one factor.

Studies of the Structure of Medical School Performance

Sirotkin and Whitten (1978) collected test score and performance rating for one class of students in an organ systems curriculum at Wayne State University's School of Medicine. This curriculum was very similar to the current curriculum at MSU-COM: Year 1 consisted of basic science courses; Year 2 was comprised of courses in organ systems biology which consisted of both clinical and basic science input; and Year 3 was a year of clinical clerkship training. Using canonical correlations, the authors correlated test scores and clinical performance ratings from each year of the curriculum with those from each other year. These canonical correlations showed considerable consistency performance during contiguous years (i.e., $R_{\text{Year 1, Year 2}} = .76$; $R_{2,3} = .71$) and a surprisingly strong relationship between performance in the basic science courses in Year 1 and clinical clerkship performance ratings during Year 3 ($R_{1,3} = .59$). The matrix of canonical correlations is displayed in Table 2.1.

Table 2.1

Canonical Correlations Among Performance Measures
(Adapted from Sirotkin & Whitten, 1978)

Year	1	2	3
1	1.00		
2	.76	1.00	
3	.59	.71	1.00

Markert (1978) investigated the relationship between classroom performance in the neuromuscular system at MSU-COM and student performance on carefully evaluated neurological history and physical examinations. He reported a significant canonical correlation of .46 between these two groups of measures.

Gough, Hall, and Harris (1964) conducted a large scale study of over 1,200 graduates from the University of California Medical School at San Francisco from 1951 to 1963. One aspect of their study was an investigation of the correlations among yearly medical school GPA's. The median correlations among GPA's for each of the four years and the median correlation between each yearly GPA and the four-year cumulative GPA are displayed in Table 2.2. Two interesting findings stand out. First, as would be logically expected, the median correlations between GPA's in contiguous years are the highest in the matrix. Second, as in the Sirotkin and Whitten study, the correlations between performance during the first two years (the basic science phase of the curriculum) and the second two years (the clinical clerkship phase) are surprisingly high indicating some degree of consistency of performance in basic science courses and clinical performance.

Table 2.2

Median Correlations Among Yearly GPA's
(Adapted from Gough et al., 1964)

Year	1	2	3	4
1	1.00			
2	.64	1.00		
3	.52	.64	1.00	
4	.38	.44	.64	1.00
Four-year cumulative GPA	.82	.83	.82	.74

Rhoads, Gallemore, Gianturco, and Osterhout (1974) compared the award of Dean's honors to students in both the basic science and clinical phases of the curriculum. Combining their data from the entering classes of 1962 to 1970 ($N = 728$), and calculating an odds ratio (Reynolds, 1977), it can be estimated that the odds in favor of clinical honors are 3.68 times as great for students who received basic science honors (1.60 to 1.00) as for those who did not receive basic science honors (0.44 to 1.00). The 95% confidence interval for this odds ratio is 3.26 to 4.10. Since the interval does not contain 1.00 (which would indicate equal odds of receiving clinical honors for both groups), it can be concluded that a significant positive association between clinical and basic science performance exists. The strength of association between the two types of performance can be estimated by using Yule's Q (Reynolds, 1977). The Q statistic for these data is .57. Rhoads et al. observed, however, that there were several students (26% of their total sample) who received clinical honors but who did not receive basic science honors. They therefore concluded that proficiency in the basic sciences is not the sole determiner of success in the clinical phase.

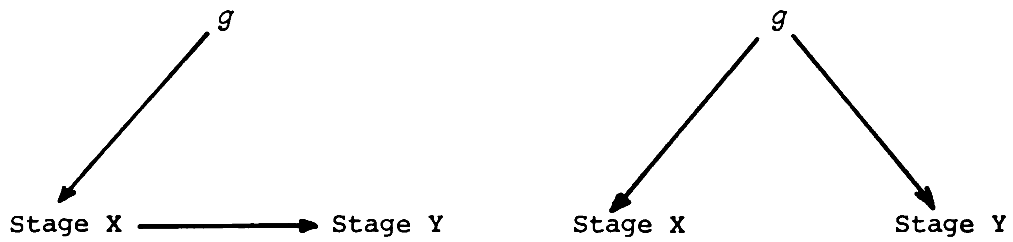
Schumacher (1964) factor analyzed medical school grades, National Board Examination scores and peer ratings of what he called "functional knowledge," diagnostic skills and skill in relating to patients for a group of interns. He reported finding a "general knowledge" factor which accounted for 44% of the total variance of the correlation matrix and which contained high loadings for medical school grades, scores on Parts 1 and 2 of the National Boards, and peer ratings of functional knowledge and diagnostic skill. Ratings of functional knowledge, diagnostic skill, and skill in patient relationships loaded on a second orthogonal factor which accounted for 9% of the total variance.

A similar finding of a general factor of clinical competence was reported by Maatsch, Downing, Sprafka, and Holmes (1978). Maatsch and his associates factor analyzed scores on multiple-choice tests of clinical and clinically-relevant basic science knowledge, patient management problems (PMP's), and ratings of simulated clinical encounters. Participating in the study were currently practicing emergency physicians, physicians in other specialties who were eligible to be certified as emergency physicians (the board eligible group), and medical students. Excluding four patient management problems which did not discriminate between medical students and physicians, all of the tests loaded on a single, general factor which accounted for 43% of the variance and 83% of the communality of the correlation matrix. Other factors which were found were a PMP format effect and a multiple choice question format effect, each of which accounted for an additional 6% of the communality.

An alternative way of looking at complex learning and performance is to conceptualize it as taking place in a hierarchical sequence. Using this conception, Gagne' (1974) has hypothesized a learning hierarchy in which learning and concomitant performance at one stage depends upon possessing the knowledge or skills which were acquired at the next lower stage. Thus, as a simple example, the learning and performance of multiplication should depend upon the knowledge of addition. Bloom and his colleagues (Bloom, 1956) developed a hierarchical taxonomy of processes involved in subject matter learning. The well-known Bloom taxonomy consists of the following stages: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. As in the Gagne' hierarchy, performance at one stage is assumed to be dependent upon the degree of the student's accomplishments at the preceding stage. Thus, for example, application of a principle is assumed to take place only after the student adequately comprehends the principle.

The simplest way of testing for a performance hierarchy is to see if performance at different levels is correlated (e.g., Gagne', 1974), or to compare the proportion of students succeeding at a stage n given success at stage *n-1*. If the stages are indeed hierarchical, the first proportion should be greater than the second proportion. Using the correlational model, the relationship between performance at two different levels of a hierarchy is schematically illustrated in Figure 2.2(a).

A serious potential deficiency of the correlational approach is suggested by the factor analytic studies of performance discussed above. That is, performance at any or all levels of the hierarchy may be influenced by the student's general level of ability (e.g., Jensen, 1969; Spearman, 1904) or knowledge (Ebel, 1969). If this is a tenable hypothesis, performances at different levels may, in the path analytic sense, be spuriously correlated because of the underlying "influence" of a general factor. This potential influence of a third variable, *g*, is schematically represented in Figure 2.2(b). In this figure, the student's previous level of background knowledge or ability (*g*) is shown as influencing his or her level of performance at stage X of the hierarchy, which, in turn, influences his or her level of performance at stage Y.



(a) Developmental model. (b) General factor model.

Figure 2.2. Two alternative causal models.

Thus, the first possibility is that background factors (i.e., the student's level general academic aptitude or background knowledge) causally influences performance at Stage X of the hierarchy which in turn causally influences performance at Stage Y (Figure 2.2[a]). Or, the student's general level of aptitude or knowledge underlies performance at both stages of the hierarchy, thus producing a spurious correlation between X and Y (Figure 2.2[b]). Expressing this latter relationship in factor analytic terms, both X and Y load on g. Hence, when g is statistically controlled, the partial correlation between X and Y, $r_{XY:g}$, should be close to zero. On the other hand, if a hierarchical or a developmental relationship exists among g, X, and Y (as shown in Figure 2.2[a]), $r_{XY:g}$ will probably be less than r_{XY} but will not disappear completely when g is controlled (Hyman, 1955).

Kropp and Stoker (1966) constructed four taxonomic-type tests designed to operationally define the six levels of the Bloom taxonomy in both science and social studies. On the basis of an analysis of both mean performance on the tests and an analysis of patterns of correlations among the tests, they concluded that the results generally supported the hypothesized hierarchical structure of the taxonomy.

Madaus, Woods, and Nuttall (1973) employed a causal model approach to test the cumulative structure of the six major levels of Bloom's taxonomy. Using multiple regression procedures to estimate the strengths of associations between performance at different levels of the taxonomy, they reanalyzed the Kropp and Stroker data. According to Madaus et al., the multiple R^2's between measures of performance at adjacent levels should be significant (indicating "direct" links between performance at adjacent levels of the hierarchy). However, the increment in R^2 after variance due to performance at intervening levels has been statistically controlled should not be significant indicating the absence of "indirect" links between levels in the hierarchy. In terms of the authors' causal model, this second finding would also support the hypothesis of the absence of the effects of other variables such as *g* which have not been included in the model. A causal model which depicts the situation described above is shown in Figure 2.3.

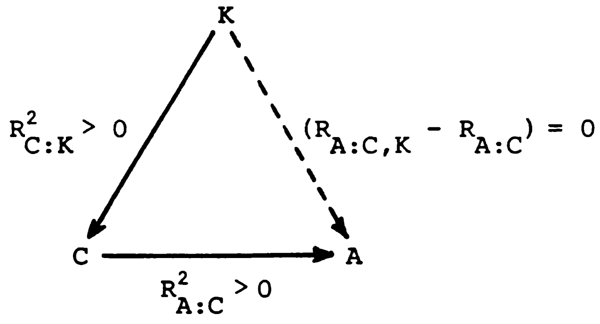


Figure 2.3. Segment of Madaus et al. causal model.

This strength of the direct link between Knowledge (K) and Comprehension (C) is estimated by $R^2_{C:K}$. Similarly, the magnitude of the direct link between Comprehension and Application (A) can be estimated by $R^2_{A:C}$. The strength of the indirect link between Knowledge and Application can be estimated by $(R^2_{A:C,K} - R^2_{A:C})$. This difference is the variance in Application which is accounted for by Knowledge when the variance due to the intervening level of Comprehension is statistically controlled. In multiple regression terms, this procedure tests the increment in the R^2 for Application when Knowledge is entered into the regression equation after Comprehension.

Testing the significance of the difference $(R^2_{A:C,K} - R^2_{A:C})$ is also statistically equivalent to testing the significance of the correlation between A and K partialing out C; that is, the partial correlation, $r_{AK:C}$. When either statistic is significantly different from zero (indicating an indirect link in hierarchy), this is a hint that a variable which has not been included in the model may be producing spurious correlations between performance at adjacent levels.

In order to test this alternative explanation, Madaus et al. performed the regression analyses again, this time controlling for students' scores on the Kit of Reference Tests for Cognitive Factors, a well-known measure of *g* developed by the Educational Testing Service. As was expected on the basis of theories of general knowledge and ability, controlling for *g* attenuated the size of the correlations between adjacent levels of the hierarchy and reduced to almost zero the strengths of all but one indirect link in the hierarchy (the link between Comprehension and Analysis). In terms of the Madaus et al. causal model, these findings indicate that the performance on one level of the hierarchy is partially dependent upon performance at the next lower level *and* partially dependent upon the student's general level of ability or knowledge. The final results of this reanalysis are shown in causal model form in Figure 2.4. The numbers in the figure are the R^2's between performance at different levels of the hierarchy.

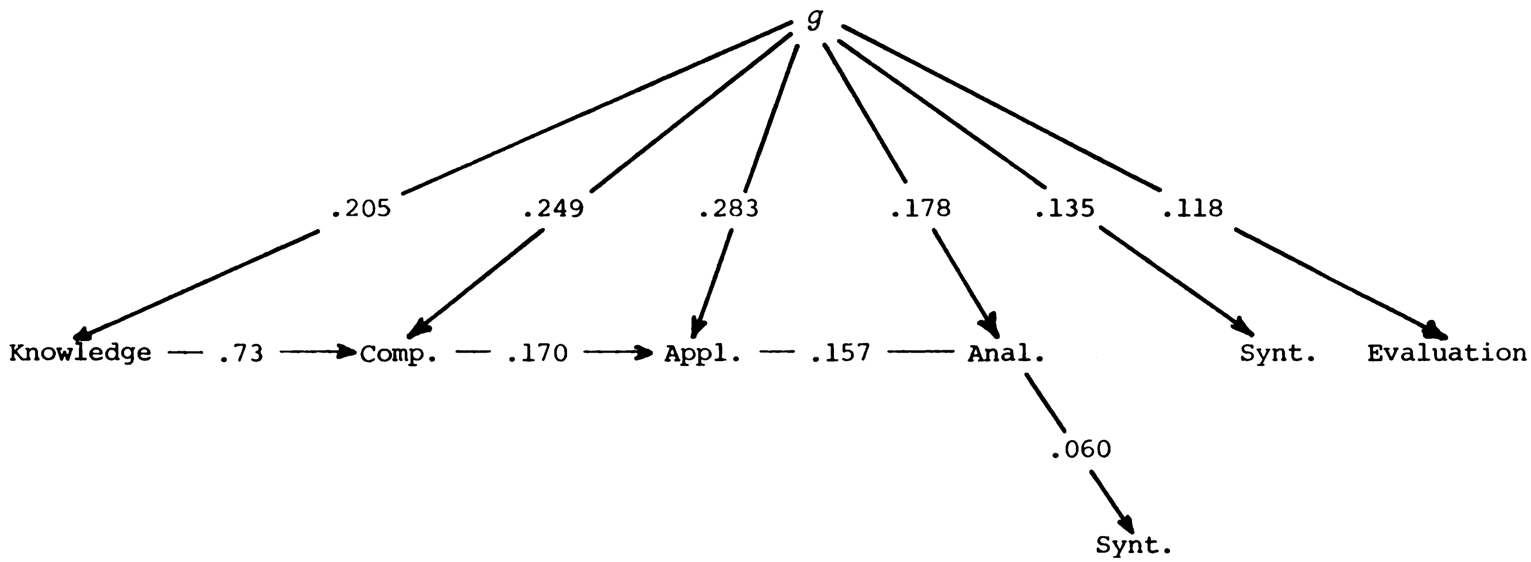


Figure 2.4. R^2's between levels of Bloom's taxonomy (adapted from Madaus et al., 1973, Figure 2).

Summary

The factor analyses of achievement and ability measures have yielded the following general findings (e.g., see Carroll, 1978; Cooley, 1976; Vernon, 1961):

1. Achievement measures tend to be moderately to highly interrelated.

2. Unrotated factor analytic solutions yield first or general factors which typically account for 30 to 50% of the total variance of the correlation matrix. For a set of measures administered to a reasonably heterogeneous group, Vernon (1961) has estimated that the general factor will account for an average of 40% of the total variance.

3. When the group of examinees or students is relatively homogeneous in ability, the average proportion of variance accounted for by the general factor decreases.

4. When the initial solution is rotated, smaller groups or clusters of measures appear (e.g., Thurstone, 1938). These group or multiple factors tend to represent cognitive and performance aptitudes or abilities.

5. When the factors are permitted to correlate (i.e., the analyst uses an oblique rotation), the multiple or group factors are typically correlated; with high correlations among the cognitive ability factors and lower correlations between cognitive and performance factors (e.g., Wolfel, 1940).

The results of studies which have investigated the consistency of medical school performance across different types of knowledge and skills have shown that such performance is relatively consistent. These findings are congruent with the results of earlier correlational studies which show achievement measures to be moderately to highly intercorrelated and the results of the factor analytic studies summarized above which support the hypothesis of a general factor of achievement or knowledge. The most critical finding reported in these studies is the surprisingly strong relationship between measures of basic science performance and measures of clinical performance (e.g., Gough et al., 1964; Maatsch et al., 1978; Sirotkin & Whitten, 1978).

An alternative way of looking at complex performance over time is to view it as stages in a hierarchy. Studies by Gagne' and his colleagues have shown that subject matter learning and concomitant performance in elementary school mathematics and science is hierarchically structured but paradoxically that instruction does not have to be sequenced in a way which is consistent with this structure for learning to take place (Gagne', 1974).

Investigations of the Taxonomy proposed by Bloom (1956) have yielded approximately similar results. That is, performance on achievement tests measuring learning in elementary and secondary school subjects has been shown to approximate the sequence proposed in the hierarchy. Consistent with the results of factor analytic studies, however, is the finding that when general ability is statistically controlled, correlations among performance at adjacent levels of the taxonomy are attenuated (Madaus et al., 1973). This result indicates the probable "influence" of general knowledge or ability on performance as well as the "influence" or what was learned at an earlier stage. The possible applicability of these hierarchical models to the description and analysis of learning processes in the MSU-COM curriculum will be discussed in the next chapter.

The Prediction of Medical School Performance

An Historical Perspective

The contributions of the Flexner Report to the curricula and admissions practices of American medical schools were discussed briefly in Chapter I. Another important influence on curriculum and, hence, on admissions policies came from the European roots of modern Americal medical education. The first U.S. medical school was established at the College of Philadelphia (later the University of Pennsylvania) in 1765 by Dr. John Morgan. Morgan, like most physicians of his time, received his medical training in Europe where the curriculum (like the typical one today) began with courses in the basic sciences and culminated in clinical clerkship training. The non-university route to an M.D. or D.O. degree was through a free-standing medical school. Flexner (1910) reported that some of these schools were barely disguised commercial trade schools whose graduates were ill-prepared for medical practice. Most proprietary schools had woefully inadequate facilities and instructors. The quality programs of the time, which were cited as exemplary by Flexner, were in university-affiliated schools whose curricula followed the European model.

As proprietary schools were refused licenses by state boards of education, and licensing and regulation, most of the medical schools which survived the impact of the Flexner Report were those allopathic, osteopathic, and homeopathic schools with adequate basic science programs. In order to admit students who would succeed in these programs, applicant selection based upon indicators of science aptitude and achievement were emphasized. As discussed in Chapter I, the most widely used indicator of science aptitude was the MCAT Science Test.

The MCAT test was originally developed in the late 1940's to equate the academic backgrounds of applicants who had attended a variety of undergraduate institutions (Erdmann, Mattson, Hutton, & Wallace, 1971). The test later became used to simply predict performance during the first two years (the basic science phase) of the traditional four-year curriculum. The original MCAT (the one used in this study) is composed of four subjects: Verbal, Quantitative, General Information, and Science. As will be discussed in the next section, the two MCAT subtests which have been most highly correlated with performance during the first two years have been the Science and Quantitative tests.

Using the age old principle that the best indicator of future performance is past performance at a similar activity, admissions committees have relied heavily upon premedical GPA as a predictor of success in medical school. Logically GPA has the following advantages as an indicator (Krupka, Elstein, Molidor, King, Parsons, & Son, 1977): It is a composite of grades earned in many courses using a variety of instructional strategies (e.g., didactic and laboratory instruction) and evaluation methods (e.g., tests, term papers, performance ratings). It can function as a relatively reliable summary estimate of performance over a long period of time (at least longer than the MCAT which samples knowledge in a variety of areas but in less than one day's testing time).

In order to get a picture of an applicant's personal qualities and to hopefully screen out undesirable personality types (e.g., the most obvious sociopaths), letters of recommendation and personal interviews are used. Applicant's letters of recommendation are almost always highly favorable, and thus are not useful in discriminating among candidates. When the applicant passes the first admissions screen (usually based upon some combination of GPA's and MCAT scores), he or she is invited to the school for a personal interview with members of the school's faculty. The applicant is typically interviewed by two faculty members who then usually rate the candidate on a series of rating scales which purport to measure personal qualities judged important in a physician (e.g., problem solving ability, decisiveness, ability to interact with others). Unless the school has a good training program the interrater reliability of the interview scores tends to be low (partially due to the low variance in the interviewers' ratings). In general, interview scores have not been significantly correlated with subsequent performance.

The Prediction of Classroom and Laboratory Performance

In the remainder of this section, representative studies concerned with the prediction of course and laboratory performance will be reviewed and discussed.

In selecting these studies, the following criteria were used: (a) the study should be fairly recent so that the results can be more readily generalized to current medical school selection problems; or (b) the study has been widely quoted in the literature in support of a certain type of admissions policy.

In most of the studies to be reviewed, medical school achievement has been operationally defined as the student's cumulative GPA during the first year, the first two years, or the entire four years of allopathic medical school. With the exception of a companion study done by the author and his associates (West, Markert, & Bernier, Note 4), the author was unable to find any investigations of the prediction of student performance in colleges of osteopathic medicine.

The most prolific investigators of the prediction of medical school performance have been Gough and his colleagues at the University of California at Berkeley (e.g., Gough, 1971, 1978; Gough, Hall, & Harris, 1963, 1964). Gough et al. (1963) studied relationships between MCAT scores, premedical GPA's, interview scores or ratings and subsequent medical school performance. Their investigation was carried out on data from over 1,200 graduates from the University of California Medical School at San Francisco between 1951 and 1962.

Gough et al. (1963) reported the following correlations between MCAT subtest scores (V = Verbal; Q = Quantitative; GI = General Information; S = Science) and first year GPA's: V = -.23 to .24 (Median = .14); Q = -.09 to .32 (Mdn = .18); GI = -.14 to .20 (Mdn = .12); S = .06 to .37 (Mdn = .28). Similar results for single classes of students have been reported by Crowder (1959): V = .14; Q = .21; GI = .09; S = .38, and by Richards and Taylor (1961): V = .16; Q = .24; GI = .09; S = .23. For ease of comparison these and other results are displayed in Table 2.3.

In spite of the intra-institutional variability reported by Gough et al., the two MCAT subtests which show the highest correlations with first-year performance across institutions are the Science and Quantitative tests. It is not surprising that these two tests (especially MCAT Science) have been among the most highly weighted criteria in the applicant selection process. The predictive validity of the MCAT Science Test was further confirmed by Gough (1978) who found the following median correlations between MCAT Science scores and yearly GPA's for the sample of University of California graduates described above: Year 1 = .28; Year 2 = .22; Year 3 = .04; Year 4 = -.03.

Buehler and Trainer (1962) analyzed the differences in scores on predictor variables for students graduating in the top 10% (22 students) and the bottom 20% (25 students) of their medical school classes. Data from six

Table 2.3

Correlations Between MCAT Scores and Medical School and Post-Graduate Performance

Study	Criterion	MCAT Verbal	MCAT Quantitative	MCAT General	MCAT Science
Gough et al. (1963)	First year GPA	.14	.18	.12	.28
Crowder (1959)	First year GPA	.14	.21	.09	.38
Richards & Taylor (1961)	First year GPA	.16	.24	.09	.23
Gough (1978)	First year GPA	--	--	--	.28
	Second year GPA	--	--	--	.22
	Third year GPA	--	--	--	.04
	Fourth year GPA	--	--	--	-.03
Johnson (1962)	Four-year GPA	--	--	--	.24
Richards, Taylor & Price (1962)	Ratings of internship performance	-.11	-.04	-.04	-.13
Howell & Vincent (1967)	Post-graduate test of medical knowledge	.47	.49	.36	.60

classes graduating between 1949 and 1954 were used. While the sample sizes are small, the study has been widely quoted. The authors found that the variable which best discriminated between the groups was MCAT Science (mean difference equalled 117 points or about 1.3 standard deviations). Other discriminating predictor variables were: premedical GPA, the other MCAT subtest scores, age of less than 25, and attendance of less than five years in a premedical program (unless the student was pursuing a graduate degree).

In a large scale and widely quoted research project, Johnson (1962) studied the relationship between two criteria of success: (a) graduation from medical school and (b) cumulative medical school performance and commonly used predictors of medical school success. Subjects included 927 applicants interviewed for admission between 1956 and 1960, 399 of whom actually matriculated, 336 who entered other medical schools, and 192 who were not admitted to any medical school. Statistically significant relationships were found between academic performance and (a) MCAT Science scores ($r = .24$), (b) premedical GPA ($r = .12$), (c) premedical science GPA ($r = .19$), and (d) quality of undergraduate institution ($r = .17$) as measured by the mean MCAT score for the student's undergraduate school, and (e) amount of outside employment during premedical studies ($r = -.15$).

"Subjective" interview scores correlated only .06 with performance, while "objective" interview scores (which were not defined further) correlated .28 with performance. Similarly, variables on which *low* scores were associated with attrition were: MCAT Science, average score on the MCAT subtests, premedical GPA, and premedical science GPA. In addition, students who were older than 28 had a significantly higher attrition rate than the national rate of 10% at that time.

Using these data on attrition rates Johnson developed a prediction index. Applicants having scores below a given cut score were assigned a "-1" for that predictor variable; middle range score, a "0," and scores above a given level "+1." For example, an applicant having a MCAT Science score below 450 was assigned a "-1"; a score of between 450 and 599, a "0"; and a score of 600 or more, a "+1." Scores on 10 admission variables were transformed using the above rules and then algebraically summed. Johnson found that 63% of admitted applicants who scored totals of -4 or -5 did not graduate; and that only 4% of those scoring -1 or above failed to graduate from medical school.

Gough et al. (1963) reported the following median correlations between premedical GPA and yearly GPA's in medical school for their sample of 1951-1952 graduates: Year 1 = .22; Year 2 = .22; Year 3 = .16; Year 4 = .07. The median correlation between premedical GPA and cumulative four-year GPA was .18. Higher correlations for a single graduating class were reported by Gaier (1952): Year 1 = .41; Year 2 = .38; Year 3 = .39; Year 4 = .32. Gough (1978) found the following correlations between premedical science GPA and yearly medical school GPA's for the Gough et al. University of California sample: Year 1 = .33; Year 2 = .23; Year 3 = .11; Year 4 = .08. For ease of comparison these results are summarized in Table 2.4. The trend in the data displayed in Table 2.4 is clear. Premedical GPA typically predicts performance during the first two years of medical school (the basic science phase) but fails to predict performance during the last two years (the clinical clerkship phase) of the traditional four-year curriculum.

In contrast to most prediction studies, Best, Diekema, Fisher, and Smith (1971) used multiple criterion measures of success in medical school. In addition to looking at yearly comprehensive exam scores at the University of Illinois, the authors collected data on clinical clerkship ratings and performance of patient management problems. The predictors which displayed the highest multiple correlation with comprehensive exam performance were: premedical GPA, Science MCAT, quality of undergraduate institution, trend in premedical GPA, and quantitative MCAT.

Table 2.4

Correlations Between Premedical GPA's and Medical School Performance

	Criterion				
Study/Predictor	Year 1 GPA	Year 2 GPA	Year 3 GPA	Year 4 GPA	Four-year GPA
Gough et al. (1963)					
Premedical GPA	.22	.22	.16	.07	.18
Gaier (1952)					
Premedical GPA	.41	.38	.39	.32	--
Gough (1978)					
Premedical Science GPA	.33	.23	.11	.08	--
Johnson (1962)					
Premedical GPA	--	--	--	--	.12
Premedical Science GPA	--	--	--	--	.19

The authors reported the following multiple correlations between these predictors and yearly comprehensive examination performance: Year 1 = .51 (R^2 = .26); Year 2 = .42 (R^2 = .18); Year 3 = .40 (R^2 = .16); Year 4 = .39 (R^2 = .15). The best single predictor of comprehensive exam performance was GPA. Best and his associates reported an overall multiple correlation of .55 (R^2 = .30) between the multivariate combination of performance measures and MCAT Quantitative, MCAT Science, premedical GPA, and quality of undergraduate college.

Presumably in response to the social turmoil and demands of the 1960's, medical schools began to put more emphasis on noncognitive admissions criteria (e.g., evaluations of letters of recommendation, indicants of the applicant's social commitments). This change in emphasis plus the implementation of affirmative action programs has broadened the traditional acceptance pool. Theoretically, as students with lower GPA's and MCAT scores are admitted to medical schools and the variances of these variables in the pool of accepted applicants increase, their correlations with medical school performance should increase as well.

The results of a longitudinal study by Frederick McGuire empirically confirm this relationship. Using percentile ranks of premedical GPA, Science GPA, MCAT

Science, and MCAT Quantitative scores. McGuire (1977) developed a multiple regression-based index for predicting academic success. Correlations between values on the index and first year class rank at the University of California at Irvine are shown in Table 2.5. The increase in the magnitude of the correlations clearly covary with the increases in the variability of the index itself. This relationship suggests that when the range of GPA's and MCAT scores of matriculants is increased, the correlations between these variables and some criterion or criteria of academic success will increase as well.

Table 2.5

Correlations Between Regression-Based Prediction Indices and First Year Class Ranks (Adapted from F. McGuire, 1977, Table 1, p. 417)

Entering class	No. of students	Index score	
		r	S.D.
1965	85	.37	63
1966	85	.46	69
1967	63	.34	42
1968	61	.31	32
1969	62	.26	27
1970	63	.33	34
1971	66	.49	50
1972	69	.49	40
1973	58	.49	44
1974	69	.84	90

As noted above, applicants who pass the first admissions screen are invited for interviews with two or more of the school's faculty members. In some schools potential interviewers are carefully trained in interview workshops and instructed in how to structure the interview and rate the candidate on the school's interview rating form. In other institutions faculty members are asked to volunteer to be interviewers in their spare time and are given little or no training. As would be expected from other social science research, interrater reliability of the interview ratings or scores generally increases with the amount of training and the degree of structure of the interviews. In general, however, interrater reliability has not been high nor have significant relationships between interview scores and subsequent student performance been reported (e.g., Krupka et al., 1977).

Prediction of Clinical Performance

The major measures of clinical performance have been ratings of clerkship, internship, and residency performance. In situations in which the practicing physician functions as an employee of an organization (e.g., the U.S. Public Health Service, the Veterans Administration), supervisors' ratings of performance have been available. However, since most physicians are self-employed, ratings of post-graduate clinical performance have not usually been available. The new trend in research on physicians' performance on specialty board exams (such as the Maatsch et al. study discussed earlier), and the study of the medical inquiry process by Elstein, Shulman, Sprafka and others (1978) will provide additional information.

In a study conducted at the College of Human Medicine at Michigan State University, Krupka et al. (1977) studied the prediction of medical students' problem solving and empathy skills. Problem solving skills were measured by ratings of the clerkship student's clinical problem solving skills by both peers (i.e., other medical students) and by clinical faculty, multiple choice tests of clinical knowledge, and diagnostic patient management problems. Variables used to predict these problem solving criteria were the MCAT, premedical GPA, the Watson-Glaser Critical Thinking Appraisal Test, the State-Trait Anxiety Inventory, the IPAT Anxiety Scale, and the Study Habits Inventory. A separate multiple regression was done for each dependent variable. MCAT scores and GPA were entered on the first step and the other predictors on the second step of the regression. Contrary to the findings about to be discussed, moderate and statistically significant multiple correlations were found between MCAT scores and premedical GPA, on the one hand, and the separate problem solving criteria, on the other. After the other predictors were entered into the regression equation on the second step, multiple R's of above .4 were reported for the following criteria: faculty ratings of problem solving, patient management problems, and scores on the multiple choice exams.

Similar regressions were carried out on peer and faculty ratings of students' empathy skills. Using only MCAT and GPA as predictors yielded multiple R's of .358 ($R^2 = .128$) and .593 ($R^2 = .352$) for peer and faculty ratings of empathy skills, respectively. When scales which were developed by the authors to measure empathy skills were added to the regressions, the multiple R^2's increased to .335 and .783, respectively.

In contrast to these findings most studies of clinical clerkship performance have reported little or no relationship between MCAT scores and ratings of clinical clerkship performance. For example, Best et al. (1971) reported a multiple correlation of .32 ($R^2 = .10$) between the set of predictors discussed earlier and ratings of clerkship performance. Gough et al. (1963) found non-significant bivariate relationships between MCAT scores and premedical GPA's, on the one hand, and clinical clerkship performance, on the other.

Richards, Taylor and Price (1962) analyzed the relationship between interns' MCAT scores and supervising physicians' ratings of internship performance. They found the following correlations: MCAT-V = -.11; Q = -.04; GI = -.04; S = -.13. These researchers did, however, report the following significant correlations between ratings of internship performance and the following yearly medical school GPA's: Year 1 = .21; Year 2 = .24; Year 3 = .45. Similar relationships were reported by Kegel-Flom (1975). Cumulative GPA was significantly correlated with supervisor ratings (.32), self ratings (.46), and peer ratings (.35) of internship performance for 110 graduates of the University of California Medical School at San Francisco. In contrast to these findings, Korman and Stubblefield (1971) found no relationship between medical school grades and interns' clinical performance.

Howell (1966) dichotomized supervisors' comments about the performance of 312 federally employed physicians into "high" and "low" performance ratings. She found no significant differences between the performances of either group on the four MCAT subtests. Howell and Vincent (1967) in a study of U.S. Public Health Service physicians found the following significant correlations between MCAT scores and scores on written examinations of medical knowledge: V = .47; Q = .49; GI = .36; S = .60. However, no significant correlations were found between MCAT scores and scores on the clinical medicine portion of this same written exam.

Their most surprising finding (and the one which is most widely quoted by critics of the MCAT) was the report of a greater than chance number of low but significant negative correlations between MCAT scores (especially on the Verbal and General Information subtests) and supervisors' ratings of clinical performance. Bartlett (1967) followed 49 medical school graduates through medical school and into the beginnings of their professional careers. He failed to find any significant differences in career performance between high and low MCAT scorers.

Wingard and Williamson (1973) reviewed 27 studies of the relationships between professional or graduate school grades and subsequent professional performance in medicine and other fields. No consistently strong relationships were found between grades and post-graduate performance in any of these fields.

The most consistent patterns in these findings are: (a) the most successful premedical predictor of clinical performance is premedical GPA (studies have not demonstrated the predictive validity of MCAT scores) and (b) predictors of clinical performance correlate most highly with those criteria of clinical performance which have the closest temporal relationship with the predictor. That is, premedical grades and MCAT scores are moderately correlated with problem solving and empathy skills which have been assessed during the clerkship years (Krupka et al., 1977) but are not correlated with internship performance (which is measured one or more years later). Medical school grades are significantly related to intership performance (e.g., Richards et al., 1962) but are not significantly related to post-graduate professional performance (Wingard & Williamson, 1973). Other studies have reported no relationships between cognitive predictors and clinical performance (e.g., Korman & Stubbelfield, 1971).

Some of the possible explanations for the lack of strong correlations between predictors and criteria of clinical performance are the following:

1. Lack of well defined criteria.

2. The low reliabilities of clinical performance ratings.

3. The lack of variance in the ratings.

4. The case specificity of the ratings. For example, Elstein et al. (1978) reported low correlations among the ratings of the performance of practicing physicians on a variety of standardized, simulated cases.

In addition to these possible reasons, most of the studies which have been reviewed in this and the previous section have used bivariate correlational techniques to estimate the relationships which were investigated.
Two more appropriate techniques for investigation of

relationships involving multiple predictors and/or criteria would have been multiple regression and canonical correlation. These methods would have yielded more statistically powerful estimates of the relationships among sets of predictor and/or criterion variables. Similarly, the reliabilities of the measures of clinical performance may have been improved upon by forming linear combinations of these individual measures (e.g., Nunnally, 1967).

Summary

Of the principally used variables for selecting students for admission to medical school, the ones which show the strongest relationships with classroom and laboratory performance are: premedical GPA, MCAT Science and Quantitative scores, and the quality of the applicant's undergraduate institution. Two variables which are negatively related to medical school performance during the basic science phase are the student's age and extent of previous employment. While statistically significant, the correlations between these predictors and first and second year grades in medical school (i.e., performance during the basic science phase of the curriculum) are generally relatively low in magnitude, and tend to be unstable from year to year at the same school. Proponents of the continued use of these selection variables in admissions decision making attribute the low to moderate magnitudes of the correlations to the restriction in range in the selection variables. This hypothesis has been supported to some extent by the results of the study by McGuire (1977). As McGuire's medical school relaxed its admission criteria, the range in these selection variables increased and their multiple correlation with first-year academic performance increased concomitantly.

The selection variable which correlates most highly with clinical performance in medical school is premedical GPA. While MCAT scores correlate surprisingly well with performance on *written* tests of medical knowledge taken after graduation (e.g., Howell and Vincent, 1967), they do not generally correlate with ratings of clerkship, internship, or post-graduate clinical performance. The most consistent pattern in the findings on the prediction of clinical performance is that cognitive predictors of clinical performance correlate most highly with those criterion measures which have the closest temperal relationships with that predictor. No consistent relationships have been found with non-cognitive predictors of clinical performance.

Summary and Discussion

Schumacher (1964) factor analyzed written tests of medical knowledge and ratings of clinical performance. Both types of measures loaded on the first or general factor which accounted for 44% of the total variance. Ratings of doctor-patient relationships formed the principal loadings on the second (orthogonal) factor which accounted for 9% of the total variance. Maatsch et al. (1978) also factor analyzed objective measures of clinically relevant basic science knowledge, clinical knowledge and ratings of diagnosis and case management. They reported a single factor solution in which the general factor accounted for 43% of the total variance. Non-factor analytic studies have demonstrated similar consistency in performance across basic science and clinical subject matter areas, and knowledge and performance domains. Both Gough et al. (1964) and Sirotkin and Whitten (1978) reported moderate to sizeable correlations among performance in basic science courses, clinical medicine courses, and ratings of clinical clerkship performance.

These findings are similar to those which have been reported in analysis of achievement and ability variables in other populations: First, achievement or ability measures tend to be moderately to highly interrelated. Second, when these measures are factor analyzed, unrotated factor analytic solutions typically yield first or general factor on which most of the measures load and which account for about 40% of the total variance of the correlation matrix (Vernon, 1961).

In the majority of the studies which have included both objective measures of performance and ratings of (mainly) clinical performance, however, the observed correlations among the objective measures have been higher than the correlations between the objective measures and the ratings. There are three probable reasons for this discrepancy: (a) objective measures are generally more reliable than ratings; (b) written tests generally have larger variances; and (c) performance on the written examinations may simply be due to test wiseness, ability to memorize, better study habits, or other reasons which some would argue are not truly related to being a competent physician. These critics would argue, therefore, that objective tests are "tapping" these "irrelevant" qualities rather than important knowledge.

An alternate view of the structure of complex performance has been offered by educational psychologists. Gagne' (1974) and Bloom (1956) have proposed hierarchical models of performance in which performance at one level is hypothesized to be dependent upon the acquisition of the knowledge or skills which comprise performance at lower levels rather than a student's general level of ability.

Research by Gagne' and others on mathematics and science learning has demonstrated that the performance structures in these subjects consist of a series of hierarchically ordered skills and that performance of these skills was correlated (Gagne', 1974). Similar findings have resulted when Bloom's taxonomy has been studied (e.g., Kropp and Stoker, 1966; Madaus et al., 1973). It is still possible, however, that this consistency of performance across levels is due to the student's general level of knowledge (Ebel, 1969), or ability (e.g., Jensen, 1969; Spearman, 1904). When Madaus et al. (1973), statistically controlled for a measure of general ability in their analysis, the correlations between performance at non-adjacent levels of the hierarchy virtually disappeared, and the correlations between performance at adjacent levels of the hierarchy were attenuated. Madaus and his colleagues were led to the "compromise" conclusion that performance at one level was partially due to the mastery of the learning process at lower levels and partially due to general ability.

Viewed from the perspective of theories of general knowledge or ability the consistency of performance in medical school could be attributed to the student's "general level of ability" (e.g., Maatsch et al., 1978). Looked at from the perspective of theories of learning or performance hierarchies, clinical performance is based upon knowledge and principles of medical biology and clinical medicine acquired earlier in the curriculum. Thus, students who have more thoroughly acquired these "basics" would be predicted to be more highly rated in their clinical performance. The third conclusion which represents a compromise between the first two is that performance at all "levels" of the curriculum is a joint function of knowledge, skills, and principles acquired at earlier levels, *and* general level of ability.

Most modern American medical curricula still follow the European model: Two years of basic science education followed by two years of clinical training. The problem of selecting medical students has, in practical terms, been reduced to selecting students who will academically succeed in these curricula. Studies which have investigated the relationships between typically employed selection variables and later performance in medical school were reviewed. The general conclusions of this review were in accord with the conclusions of previous reviews of the literature in this area. The best predictors of performance during the basic science years of the curriculum have been: MCAT scores (especially MCAT Science), premedical GPA (especially in the sciences), and the quality of the applicant's undergraduate institution. These objective measures of the applicant's academic achievement and aptitude have been found to have low to moderate correlations with later academic performance (e.g., Best et al., 1971; Gough, 1979; Gough et al., 1963; Johnson, 1962). A multi-year study of matriculants recently done by McGuire (1977), however, demonstrated that as the variances of such objective predictor variables increased, their multiple correlation with first year academic performance increased concomitantly.

MCAT scores and premedical GPA's were not generally found to be well correlated with ratings of clinical performance during the last two years of medical school (e.g., Best et al., 1971; Gough et al., 1963), internship performance (e.g., Richards et al., 1962), or post-graduate professional performance (Howell, 1966; Howell & Vincent, 1967). On the other hand, the best predictors of clinical performance were those which had the closest temporal relationship to the clinical performance being predicted. For example, the best predictor of clinical performance during the clerkship and internship years were medical school grades during the previous years (e.g., Richards et al., 1962; Sirotkin & Whitten, 1978).

With the exception of the applicant's age and his or her reported number of hours of outside employment during undergraduate school (both of which have been found to be negatively correlated with later performance), biographical variables have not been found to be consistently related to medical school performance. Similarly, subjective ratings of applicants' personal traits after interviews with the applicant have not been reported to be significantly related to later performance.

Based upon the findings of studies reviewed in this chapter, three general research hypotheses can be offered to guide further investigation:

1. The student's performance across different areas of the curriculum (e.g., clinical and basic science) should be consistent.

2. This performance may be structured along the lines of the learning or performance hierarchies proposed by Bloom (1956) and Gagne' (1974). The exact organization of the structure would probably be different for different medical schools. However, a general structure which would be applicable to most or all schools would probably consist of at least two stages: (a) the acquisition of basic science knowledge, and (b) the application of this knowledge in clinical performance.

3. A multivariate relationship between predictors of medical school performance and the performance itself can be hypothesized to exist.

CHAPTER III

METHOD

In order to obtain data to test the hypothesized relationships discussed in Chapter I and at the conclusion of Chapter II, MSU-COM faculty members were requested to provide summary measures of student performance in the classes which they taught. These course performance measures and the medical student samples on which they were taken are described in the next sections. Also included in this chapter are a restatement of the research hypotheses to be investigated and a description of the data analysis procedures to be used.

The Sample and the Method

Academic and clinical performance data were collected for matriculants entering MSU-COM in 1974 and 1975. Reasonably complete data (i.e., grades or other measures of summary performance in at least 75% of the courses for which data were collected) were available for 84 of the 88 students matriculating in 1974 and 96 of the 99 students matriculating in 1975.

Selected preadmissions characteristics of these students are displayed in Table 3.1 (a complete list appears in Table 3.2). The relatively wide ranges on some of these variables reflect MSU-COM's commitment in its developing years to experimentation with the admission of non-traditional students. That is, the admission of a relatively high proportion of ethnic minority students, students with non-premedical academic backgrounds, and older students applying to medical school for training for a second career. The statistical benefit of these wide ranges in the admissions predictor variables is the increased probability that these predictor variables will correlate more highly with medical school performance than they would in traditional allopathic medical programs (such as those discussed in Chapter II).

As discussed in Chapter I, MSU-COM has a three-part integrated curriculum. Courses offered during the first eight terms of the curriculum are displayed in Figure 3.1. During the first two terms of the program (Unit 1) students take mainly basic science courses (e.g., anatomy, physiology, biochemistry) as well as courses concerned with introductions to physical diagnosis, osteopathic principles and practice, family, and community medicine. The remaining six terms of on-campus osteopathic medical education (Unit 2) consist of systems biology courses which include basic

Table 3.1

Selected Preadmissions Characteristics of Matriculants

Characteristic	Class 1974	Class 1975
Sample size	84	96
Age		
Mean	24.3	24.5
Standard Deviation	4.3	6.6
Range	20-42	19-40
Sex		
Male	71%	75%
Female	29%	25%
Ethnic status		
Minority	21%	17%
Majority	79%	83%
Undergraduate major		
Biological Science	55%	48%
Health related	21%	31%
Non-Biological Science	7%	8%
Other	17%	13%
Premedical GPA		
Mean	3.01	3.15
Standard Deviation	0.37	0.34
Range ($\bar{X} \pm 2$ s.d.)	2.27-3.75	2.46-3.83
Premedical Science GPA		
Mean	2.97	3.09
Standard Deviation	0.43	0.39
Range ($\bar{X} \pm 2$ s.d.)	2.11-3.83	2.31-3.87
Premedical Non-Science GPA		
Mean	3.05	3.22
Standard Deviation	0.41	0.37
Range ($\bar{X} \pm 2$ s.d.)	2.23-3.87	2.48-3.96
MCAT Verbal		
Mean	414.20	513.82
Standard Deviation	85.80	93.71
Range ($\bar{X} \pm 2$ s.d.)	323-366	326-701
MCAT Quantitative		
Mean	523.95	555.24
Standard Deviation	97.49	81.88
Range ($\bar{X} \pm 2$ s.d.)	334-720	392-719
MCAT General		
Mean	502.70	511.31
Standard Deviation	83.62	87.72
Range ($\bar{X} \pm 2$ s.d.)	336-670	336-686
MCAT Science		
Mean	511.71	521.07
Standard Deviation	96.93	90.05
Range ($\bar{X} \pm 2$ s.d.)	415-609	431-612

Table 3.2

Complete List of Preadmissions Characteristics of Matriculants

Biographic

1. Sex
2. Application-reapplication (Was the student accepted on his first or succeeding attempts?)
3. Original-alternate (Was the student selected originally or as an alternate?)
4. Age
5. Majority-minority status (Majority = Caucasian; Minority = Other)
6. Marital status (Married or not married?)
7. Military service (Was the student in the military or not?)
8. Residency (Is the student a Michigan resident or out-of-state resident?)
9. Number of schools (How many postsecondary institutions did the student attend?)

Course Work

10. Undergraduate science GPA
11. Undergraduate science credit hours
12. Undergraduate nonscience GPA
13. Undergraduate nonscience credit hours
14. Overall undergraduate GPA
15. Overall undergraduate credit hours
16. Biology GPA
17. Biology credit hours
18. Inorganic chemistry GPA
19. Inorganic chemistry credit hours
20. Organic chemistry GPA
21. Organic chemistry credit hours
22. Physics GPA
23. Physics credit hours
24. English GPA
25. English credit hours
26. Behavioral science GPA
27. Behavioral science credit hours

Medical College Admission Test (MCAT)

28. MCAT Verbal
29. MCAT Quantitative
30. MCAT General
31. MCAT Science
32. MCAT Average

Table 3.2--Continued

Other

33. Score for admission interviews
34. D.O. hospital experience (Prior to admission did the student have or not have D.O. hospital experience?)
35. D.O. relative (Does or does not the student have a D.O. relative?)
36. D.O. nonrelative (Does or does not the student have a D.O. nonrelative contact?)
37. D.O. friend (Does or does not the student have a D.O. friend?)
38. Work (Prior to admission how many hours per week did the student work?)
39. Health-related activity (Prior to admission was or was not the student involved in a health-related activity?)
40. Extra-curricular activity (Prior to admission was or was not the student involved in an extra-curricular activity?)

		Type of Course					
		Basic Science	Systems Biology	Clinical Science	Osteopathic Manipulative Therapy (OMT)	Family Medicine Preceptorships	Community Medicine
Year 1	Unit 1: Term 1	Physiology Anatomy Histology Biochemistry		Physical Examination			CM 510
Year 1	Term 2	Pharmacology Microbiology Pathology		Physical Diagnosis	OMT 1	FM 632	CM 511
Year 1	Unit 2: Term 3	Clinical Pharmacology	Hematopoetic Integumentary Endocrinology	Clinical Science 3	OMT 2	FM 642	CM 512
Year 1	Term 4		Neurology	Clinical Science 4	OMT 3	FM 652	CM 513
Year 2	Term 5		Cardiovascular	Clinical Science 5	OMT 4	FM 662	CM 514
Year 2	Term 6		Respiratory Urinary	Clinical Science 6	OMT 5	FM 672	CM 515
Year 2	Term 7		Gastro-intestinal	Clinical Science 7	OMT 6	FM 682	Psycho-pathology
Year 2	Term 8		Growth & Development Orthopedics	Clinical Science 8	OMT 7	FM 692	Psycho-pathology

Figure 3.1 MSU-COM curriculum by type of course.

and clinical science material related to the major organ systems of the body (e.g., the neuro-muscular system, the cardiovascular system). During Unit 2 students also continue their training in diagnosis, case management, osteopathic diagnosis and treatment, family and community medicine, and psychiatry. The third calendar year of the program (Unit 3) consists of traditional clerkships in community hospitals and ambulatory clinics. During this year the student's time is divided among clinical clerkships in internal medicine, surgery, pediatrics, family medicine, and other divisions of clinical medicine.

The number and types of courses for which data were collected are shown in Table 3.3. Data were available for 81% of the courses taken by students in the Class of 1974 and for 47% of the courses taken by students in the Class of 1975. The course performance measures reported by the faculty members represent summaries of the student's performance in that course which were used to determine the student's Pass/No Pass grade. These course performance measures are described in Tables 3.4 and 3.5. Most of the course performance measures in the didactic courses are weighted averages of the student's performance on objective (i.e., multiple-choice and true-false) exams. Some of the course performance measures from the basic science courses are also composed of ratings of clinical

Table 3.3

Achievement Data for Classes of 1974 and 1975

Type of course	No. of courses in MSU-COM curriculum	No. of courses for which data were collected	
		Class of 1974	Class of 1975
Basic science	9	8	8
Community medicine	8	7	3
Clinical science	8	4	2
Systems biology	8	7	4
Osteopathic manipulation therapy	7	6	3
Family practice-preceptorship	7	6	2
Total	47	38	22

laboratory skills, and some of the course performance measures from the systems biology courses also include ratings of clinical skills. The evaluation of diagnostic and case management skills in the systems courses is, however, mainly done by describing a case on paper and asking multiple-choice questions about it. Similarly, paper patient management problems are sometimes used in the small group discussion sections of the systems course.

Course performance measures from courses in physical examination, and clinical science (now called the Comprehensive Patient Evaluation sequence) are based upon faculty members' ratings of physical examination and diagnostic

Table 3.4

Course Performance Measures--Class of 1974

Course	Course identifier	Composition of course performance measure
Physiology	PSL 500	Objective
Histology	ANT 560	Objective
Anatomy	ANT 565	Objective + practical exam
Biochemistry	BCH 501	Objective
Pharmacology	PHM 520	Objective
Clinical Pharmacology	PHM 521	Objective
Microbiology	MPH 521	Objective + lab. skills
Pathology	PTH 502	Objective + lab. skills
Hematopoetic System	Hemato	Objective
Neurology System	Neuro	Objective + video cases
Cardiovascular System	CV	Objective + EKG reading
Respiratory System	Respir	Objective
Urinary System	Urinary	Objective
Gastrointestinal	GI	Objective
Growth and Development System	GD	Objective
Orthopedics System	Ortho	Objective
Physical Examination	Phyex 1	Objective + examination skills
Physical Diagnosis	Phyex 2	Objective + examination skills
Clinical Science	Clsci 6 Clsci 7	Objective + examination skills
Osteopathic Diagnosis and Manipulative Therapy (OMT)	OMT 1 OMT 2 OMT 3 OMT 4 OMT 5 OMT 6	Diagnostic and treatment skills exams + objective exams
Community Medicine:		
Medicine and Society	CM 510	Objective
Biostatistics	CM 512	
Medical Jurisprudence	CM 513	
Health Care Delivery I	CM 514	
Health Care Delivery II	CM 515	
Psychopathology	CM 5167	
Family Medicine	FM 632-692	Physicians' ratings

Table 3.5

Course Performance Measures--Class of 1975

Course	Course identifier	Composition of course performance measure
Physiology	PSL 500	Objective
Histology	ANT 560	Objective
Anatomy	ANT 565	Objective + practical exam
Biochemistry	BCH 501	Objective
Pharmacology	PHM 520	Objective
Clinical Pharmacology	PHM 521	Objective
Microbiology	MPH 521	Objective + lab skills
Pathology	PTH 502	Objective + lab skills
Hematopoetic System	Hemato	Objective
Neurology System	Neuro	Objective + video cases
Integumentary System	Integ	Objective
Endocrine System	Endoc	Objective
Physical Examination	Phyex 1	Objective + examination skills
Physical Diagnosis	Phyex 2	Objective + examination skills
Osteopathic Diagnosis and Manipulative Therapy (OMT)	OMT 1, OMT 2, OMT 3	Diagnostic and treatment skills exam + objective exams
Community Medicine:		
Medicine and Society	CM 510	Objective
Biostatistics	CM 512	Objective
Medical Jurisprudence	CM 513	Objective
Family Medicine	FM 632-642	Physicians' ratings

skills as well as objective tests of knowledge of these skills. Similarly, measures from all of the courses in the osteopathic manipulative therapy (OMT) sequence consist of ratings of the student's osteopathic examination, diagnosis, and manipulative therapy skills as well as objective tests on comprehension of the basic science knowledge and clinical principles underlying these techniques. Scores from the Family Medicine preceptorships (FM 632 to FM 692) are unweighted averages of preceptors' ratings of the student's clinical skills on ten Likert scale items.

All of the objective tests and clinical rating scales used to measure students' achievement and performance were locally constructed by MSU-COM faculty members. The internal consistency reliabilities of the single tests and other instruments which were combined to form the course performance measures are typical of instructor-made tests (i.e., in the range between .60 to .90). However, since each course performance measure was normally a composite of at least two measures, the reliabilities of the composite course performance measures are higher than the reliabilities of the single measures of which they are comprised (e.g., Nunnaly, 1967). Due to the lack (at that time) of a strong faculty development program, the reliabilities of ratings of clinical skills are not as adequate. This is especially true of the ratings of student clinical performance during the Family Medicine preceptorships as these ratings were made by off-campus, volunteer physicians. This is also partially true of the ratings of clinical skills during the OMT sequence. However, since the course performance measures for these courses are also composed of written objective tests, these course performance measures are moderately reliable.

In Figure 3.2 (adapted from Tinning, Note 2, Figure 2) the MSU-COM curriculum is schematically illustrated. The percentages in each block indicate the percentage of instructional hours devoted to each major topic during the specified time period. These percentages were arrived at through an hour-by-hour analysis of COM course protocols. As shown in Figure 3.2, as the curriculum progresses from Year 1 to Year 3, the amount of clinical instruction gradually increases. The osteopathic physicians teaching in later systems biology and clinical skills courses assume that students have mastered the basic medical biology, and knowledge of clinical principles and skills presented in earlier courses. Hence, these clinical instructors teach more sophisticated principles of diagnosis and patient management which are based upon the rudimentary concepts, skills, and vocabulary presented in earlier courses. Concomitant with these curricular changes, is a change in the

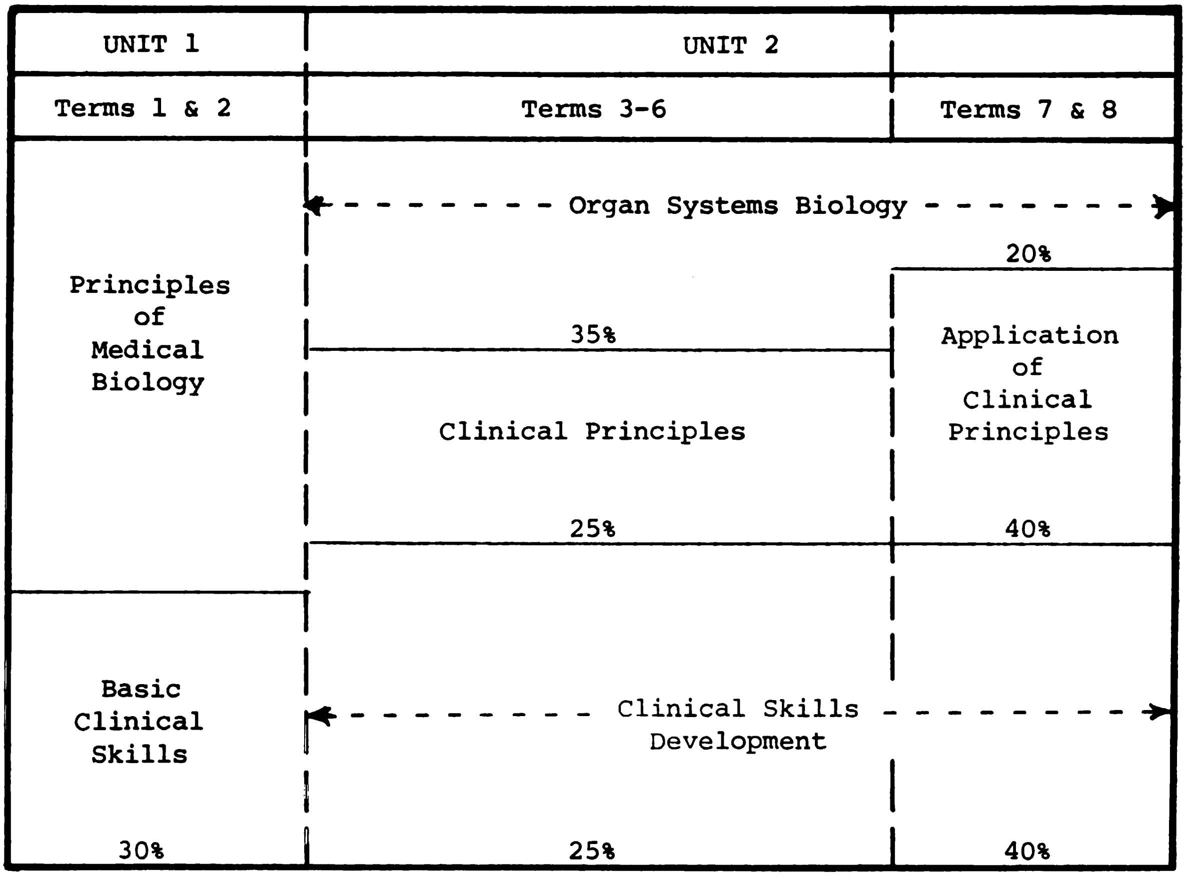


Figure 3.2. Percentage of student contact hours devoted to basic science and clinical instruction (adapted from Tinning, Note 2, Figure 2).

mode of evaluation from an assessment of the recall of specific facts and concepts to the assessment of the application of this knowledge to diagnostic and treatment situations.

Tinning, Taylor, and West (Note 3) analyzed the types and numbers of clinical experiences in MSU-COM courses offered during 1974 and 1975 (i.e., the first two years of the program for students in the Class of 1974). These researchers looked at clinical content of the courses along two dimensions:

1. What the student does during the learning session, i.e.,

   a. Acquisition of factual material;

   b. Acquisition of diagnostic data on real or simulated patients;

   c. Formulation of diagnostic hypotheses; and

   d. Treatment of real or simulated patients.

2. The mode of instruction

   a. Didactic presentations;

   b. Demonstrations of diagnostic or treatment procedures; and

   c. Hands-on performance of diagnostic or treatment procedures.

Tinning et al. concluded that these two dimensions interacted in the following way in the curriculum:

> It was found that courses at the beginning of the program stress the acquisition of factual material through large and small group didactic presentations which *describe* clinical procedures and techniques whereas, courses which occur later in the curriculum shift the instructional mode to demonstration of clinical procedures and the performance of these procedures by the students themselves; while at the same time, the student performance mode is shifting to an increasing emphasis on data *utilization* and *treatment*. (p. 2)

Using the dimensions discussed in previous paragraphs, it is possible to break down the COM curriculum into roughly three stages:

1. Acquisition of basic facts, principles, vocabulary, and skills during the first two terms (Unit 1).

2. Acquisition of principles of clinical medicine, additional basic science knowledge, and clinical skills during the first part of Unit 2.

3. Acquisition and application of more sophisticated principles and diagnosis and treatment during the later systems biology courses in Unit 2.

Using both the hour-by-hour analysis of course content and the Tinning et al. analysis of the instructional dimensions and content of clinical courses, the courses for which course performance measures are available are classified according to the structure proposed above. This classification is shown in Figure 3.3.

Course or type of course	Basic knowledge	Clinical principles	Clinical application
Physiology	+		
Anatomy	+		
Histology	+		
Biochemistry	+		
Pharmacology	+		
Microbiology	+		
Pathology	+		
Clinical Pharmacology		+	
Hematopoetic		+	
Neurology		+	
Cardiovascular		+	
Respiratory		+	
Urinary		+	
Gastrointestinal			+
Growth & Development			+
Orthopedics			+
Psychopathology			+
Physical Examination	+		
Physical Diagnosis	+		
Clinical Science			+
OMT 1, 2	+		
OMT 3, 4		+	
OMT 4, 6			+
Preceptorships			+

Figure 3.3. Hypothesized three-factor performance structure.

An alternate structure which is based upon course content alone is the following:

1. Basic science courses;
2. Early systems biology courses;
3. Later systems biology courses; and
4. Clinical skills courses consisting of all physical examinations, clinical science, osteopathic manipulative therapy courses, and family medicine preceptorships.

This hypothesized structure (shown in Figure 3.4) partially ignores both the temporal dimension and the hierarchical building upon previously learned knowledge and skills which were incorporated into the first model. That is, the "clinical skills" category has been formed by combining all courses concerned with teaching clinical skills regardless of when they were offered. One advantage of this model, however, is that it allows for a strong test of the hypothesis that performance of clinical skills is unrelated to academic performance in basic and clinical science.

A test of the two-factor, clinical and basic science performance model can be made by using the performances structure shown in Figure 3.5. A test of the single factor performance structure obviously implies that all course performance measures should load on a single factor, and thus is not illustrated.

Course or type of course	Basic knowledge	Clinical principles	Clinical application	Clinical skills
Physiology	+			
Anatomy	+			
Histology	+			
Biochemistry	+			
Pharmacology	+			
Microbiology	+			
Pathology	+			
Clinical Pharmacology		+		
Hematopoetic		+		
Neurology		+		
Cardiovascular		+		
Respiratory		+		
Urinary		+		
Gastrointestinal			+	
Growth & Development			+	
Orthopedics			+	
Psychopathology			+	
Clinical Science				+
OMT				+
Preceptorships				+

Figure 3.4. Hypothesized four-factor performance structure.

Course or type of course	Basic science	Clinical
Physiology	+	
Anatomy	+	
Histology	+	
Biochemistry	+	
Pharmacology	+	
Microbiology	+	
Pathology	+	
Clinical Pharmacology		+
Hematopoetic		+
Neurology		+
Cardiovascular		+
Respiratory		+
Urinary		+
Gastrointestinal		+
Growth & Development		+
Orthopedics		+
Psychopathology		+
Physical Examination		+
Physical Diagnosis		+
Clinical Science		+
OMT		+
Preceptorships		+

Figure 3.5. Hypothesized two-factor performance structure.

When several measures of performance are available, a research method which can provide a great deal of information about their latent or underlying structure is covariance structure analysis (e.g., Jöreskog, 1974; Wiley, Schmidt, & Bramble, 1973). A subset of covariance structure analysis techniques is confirmatory factor analysis. As briefly discussed in Chapter I, confirmatory factor analysis allows the researcher to test specific hypotheses about the latent structure of a set of data and to estimate the relationships among the hypothesized latent factors.

One of the simplest procedures which can be done with this very general and flexible technique is to test for what Thurstone called "simple structure." That is, to test the plausibility of an hypothesized model in which the variables are assumed to have large loadings on one or at the most two factors and loadings of zero on the other factors. Using confirmatory factor analysis, the analyst would restrict a variable's loadings on certain factors to be zero and have the program estimate the remaining loadings and the correlations among the hypothesized factors. The COFAMM program for confirmatory factor analysis developed by Sörbom and Jöreskog (1976) provides maximum likelihood estimates and standard errors of these estimates for all of the parameters which are estimated by the model. These maximum likelihood estimates are the most probable estimates of the parameters given both the hypothesized model and the observed data.

Conventional (or exploratory) factor analysis yields an unrestricted estimate of the factor pattern matrix (i.e., the matrix of factor loadings). However, as discussed in Chapter I, the major disadvantage of this approach is that the solutions are not uniquely identified. That is, when an orthogonal rotation is performed in order to make the results more interpretable, there are a theoretically infinite number of orthogonal rotations which will result in different but equally mathematically legitimate factor pattern matrices and, hence, potentially different conclusions about the underlying structure of the data (Lawley & Maxwell, 1971). The major advantage of confirmatory factor analysis on the other hand is that when some of the parameters of the model are judiciously set in advance, the COFAMM program will probably yield a uniquely identified solution. Rules of thumb which are necessary but not sufficient for achieving a uniquely identified solution will be discussed below.

The general statistical model for factor analysis is (Jöreskog & Lawley, 1968, Eq. 1):

$$x = \Lambda f + e,$$

where x is a vector of p measures, f is a vector of k common factors, e is a vector of p residuals which represent the combined effect of specific factors and measurement error (i.e., the "unique variances" of the measures), and Λ is a p x k matrix of factor loadings. The residuals e are assumed to be normally and independently distributed, and to have a mean vector of 0. They are also assumed to be uncorrelated with each other and with the common factors. The dispersion or covariance matrices of f, e, and x can be represented as Φ, Ψ, and Σ. Since the residuals are assumed to be uncorrelated, Ψ is assumed to be a diagonal matrix whose diagonal elements are the estimates of the unique variances associated with each of the measures. In addition, it can also be assumed without loss of generality that the common factors have unit variances so that the diagonal elements of Φ can be specified as unities.

Given these assumptions, the expected covariance or correlation matrix (if the measures have been standardized) given an hypothesized model can be represented as (Jöreskog & Lawley, 1968, Eq. 2):

$$\Sigma = \Lambda\Phi\Lambda' + \Psi,$$

where to review:

Σ = The expected covariance matrix given a hypothesized model;

Λ = The factor pattern matrix;

Φ = A matrix containing correlations among the latent factors; and

Ψ = A matrix whose diagonal elements are the specific variances associated with each of the measures.

A necessary but unfortunately not a sufficient condition for achieving a uniquely identified solution is that at least k^2 elements of the Λ and Φ matrices should be fixed (where k = the number of hypothesized factors). As mentioned above, it can be assumed without loss of generality that the variances of the k factors are 1.00's. Hence, the k diagonal elements of the Φ matrix can be fixed to unities. In addition, fixing at least k-1 elements (i.e., factor loadings) in each column of the factor pattern matrix to zero will usually cause the solution to be uniquely identified (Jöreskog & Lawley, 1968). (After two preliminary runs using COFAMM on the data for this study, the present author found that failure to fix the diagonal elements of Φ to unities resulted in a program diagnostic that one of the diagonal elements of this matrix was not uniquely identified.)

Using the COFAMM program, models derived from the four performance structures hypothesized above can be tested. That is, the COFAMM program can be constrained to yield the following solutions:

1. A one-factor model.

2. A two-factor model in which the correlation between the factors will be estimated by the program.

3. A three or more factor model.

The goodness of fit of the expected covariance matrix to the observed covariance matrix S can be tested by using the likelihood ratio chi-square statistic. When the chi-square value is low relative to its degrees of freedom, the model can be said to "fit" the data. The simplest criterion for determining the goodness of fit, therefore, is to compare the chi-square test statistic to its expectation (i.e., its degrees of freedom). If the chi-square is less than its degrees of freedom (i.e., is non-significant), the model can be said to fit the data. However, the problem in the present study as well as in many or most applications of confirmatory factor analysis is to choose the most appropriate model from a number of hypothesized models. In this case, Jöreskog (1974) has recommended the following heuristic strategy. Compute the ratio of each chi-square to its degrees of freedom, i.e.,

$$\chi_i^2/df_i$$

where $_i$ is the test statistic for the i th model and df_i equals its degrees of freedom. Also taking into consideration the theoretical aspects underlying each model, select the model which yields the lowest of these ratios.

Two additional methods have been suggested for heuristically determining how much additional information is yielded by a more complex model (e.g., a model with more factors) over a simpler model. Since the chi-square statistics for a pair of models are additive, the analyst can test the significance of the decrease in the chi-square associated with the more complex model by testing the difference in the two chi-squares referenced to the difference in their degrees of freedom, i.e.,

$$(\chi_s^2 - \chi_c^2) \sim \chi^2_{(df_s - df_c)}$$

where χ_s^2 and χ_c^2 are the chi-squares associated with the simpler and more complex models, respectively, and df_s and df_c are their respective degrees of freedom. This test is analogous to testing the significance of the increment in R^2 (and, thus, the increase in "fit") of a regression model when an additional variable has been added to the regression equation.

A descriptive statistic which is analogous to a reliability coefficient has been proposed by Tucker and Lewis (1973). R. Burt has provided a useful explanation of this technique. The following formulas and explanations of them are adapted from Burt (1973, pp. 148-150). The sum of the squared covariance not explained by the model can be computed as

$$M = \frac{\chi^2/N}{df}$$

where χ^2 is the likelihood ratio statistic associated with the hypothesized model; N is the sample size; and df is the degrees of freedom associated with the model. Similarly, the sum of the squared covariances which are able to be explained by the model can be computed as

$$M_o = \frac{\Sigma c_{ij}^2}{df_o} = \frac{\Sigma c_{ij}^2}{r\ (r+1)/2}$$

$(i < j)$

where c_{ij} is the squared covariance between measures i and j; and r is the number of rows in the covariance matrix. Thus, the numerator of M_o is simply the sum of the squared elements below the diagonal in the observed covariance matrix, S. The expected value of the sum of the squared covariances not explained by the proposed model involving the k hypothesized factors is

$$E(M_k) = 1/N$$

Combining the above information, the Tucker and Lewis reliability coefficient can be expressed as (Burt, 1973, p. 150):

$$\hat{\rho}_k = \frac{M_o - M_k}{M_o - E(M_k)} = \frac{\text{the amount of covariation explained by a proposed structure}}{\text{the amount of covariation available to be explained by a proposed structure}}.$$

According to Burt, and Tucker and Lewis, a small value of this statistic is an indication that the proposed model is inadequate to explain the covariation among the observed variables, and that more hypothesized latent variables or factors are required. As more factors are added to the model, the value of $\hat{\rho}$ will increase asymptotically. The maximum value of the statistic is 1.00 which indicates that the proposed model fits the data perfectly.

One of the difficulties with the likelihood ratio chi-square is its sensitivity to large sample sizes. Hence using the probability level of the chi-square as the criterion for deciding whether or not the proposed model fits the data may lead the analyst to accept solutions with one or more theoretically meaningless factors in order to lower the *p*-value of chi-square to an "acceptable" level. In contrast, the Tucker-Lewis statistic is not as sensitive to sample size. The value of this property of their

their statistic is illustrated in its application to Harman's classic physical measurements example (Harman, 1967). Eight physical variables were measured on 305 girls. When the intercorrelations among the measures were originally analyzed using exploratory factor analysis, two factors, "lankiness" and "stockiness," were identified. When the same data were reanalyzed using maximum likelihood factor analysis, the likelihood ratio chi-square associated with the two-factor model had a probability level of less than .001; yet the $\hat{\rho}_k$ for the two-factor solution was .934. When a third factor was added, the significance level was still less than .01 and $\hat{\rho}$ increased to .975 indicating only a slight increase in the amount of covariation accounted by the addition of the third factor to the model. When a fourth factor was added, the *p*-value associated with the chi-square statistic was .23 and $\hat{\rho}_k$ was calculated to be .994 (Tucker & Lewis, 1973). Tucker and Lewis cite Harmon's comment on the use of the likelihood ratio chi-square statistic as the "sole arbitor" of deciding how many factors to extract (Harman, 1967):

> This example illustrates the general principle that one tends to underestimate the number of factors that are statistically significant. For twenty years, two factors had been considered adequate, but statistically two factors do *not* adequately account for the observed correlations based on a random sample of 305 girls. However, the third factor (whose total

> contribution to the variance ranges from 2 per cent to 5 per cent for the different solutions) has little "practical significance," and certainly a fourth factor would have no practical value. (p. 229)

Echoing Harman's concerns, leading proponents of the use of maximum likelihood confirmatory factor analysis have emphasized that judgments about the adequacy of the solutions yielded by the technique must be based upon the theory and hypotheses underlying the investigation and the nature of the data being analyzed as well as on statistical criteria (e.g., Jöreskog, 1969; Lawley & Maxwell, 1971).

In order to have external validity, the pattern of relationships among the different factors (i.e., the phi matrices) should be approximately the same from year to year. Thus, the phi matrices for both the 1974 and 1975 classes should be approximately the same given that the same variables are specified to load on the same factors in both models. In this study, the models of the performance structures described above will be estimated on the data from the Class of 1974. Once an appropriate model has been selected, the model for the Class of 1975 can be estimated by restricting the same variable to load on the same factors, and the phi matrices for both models can be compared.

Once an adequately fitting and theoretically appropriate model has been identified, composite observed scores on the factor or factors can be easily computed by summing students' standard scores on the course performance measures.

The advantages of unit-weighted, linear composites have been discussed by Dawes and Corrigan (1974), F. Schmidt (1971), and Wang and Stanley (1970). The advantage of using standard scores over raw scores in forming the composites are apparent upon examination of the variances of the course performance measures displayed in Table 3.6. As mentioned above, the course performance measures are themselves composites of scores on instructor-made measurement instruments. Unlike standardized tests, the scales of these instruments are entirely arbitrary and even change from year to year. As shown in Table 3.6, the variances of these scales for the Class of 1974 range from a low of 3.80 for the course performance measures for the Hematopoetic System to a high of 6,952.06 for Microbiology 521, for a ratio of 6,952.06/3.80 or 1,829 to 1.00. It is well known that measures with large variances will be weighted more heavily in a linear composite strictly because of their larger variances (e.g., Nunnally, 1967). Thus, using raw scores to form linear composites of the variables in this study would clearly bias the linear composite toward performance in the basic science courses which have the arbitrarily larger variances.

Table 3.6

Means and Variances of Course Performance Measures

Course		1974 Mean	1974 Variance	1975 Mean	1975 Variance
Physiology		74.40	70.32	76.55	46.19
Histology		74.28	190.90	71.82	112.97
Anatomy		372.69	909.37	396.68	975.50
Biochemistry		78.13	101.71	72.40	75.09
Pharmacology		79.69	26.24	81.52	23.85
Microbiology		502.222	6,952.06	82.73	27.26
Pathology		56.28	50.28	62.34	27.64
Clinical Pharmacology		77.35	25.65	78.83	22.21
Hematopoetic		28.21	3.80	20.39	3.35
Integumentary		--	--	48.18	13.54
Endocrine		--	--	41.41	11.22
Neurology		50.60	46.28	50.17	45.93
Cardiovascular		106.31	56.54	--	--
Respiratory		226.20	299.46	--	--
Urinary		50.58	74.71	--	--
Gastrointestinal		246.32	409.34	--	--
Growth & Development		79.61	40.68	--	--
Orthopedics		69.00	27.01	--	--
Psychopathology		77.83	64.25	--	--
Physical Examination		88.31	34.50	211.77	157.11
Physical Diagnosis		37.25	17.18	105.91	19.01
Clinical Science 6		34.01	11.46	--	--
Clinical Science 7		33.75	6.29	--	--
Osteopathic Diagnosis & Manipulative Therapy (OMT)	OMT 1	91.19	19.24	89.33	16.89
	OMT 2	84.02	25.83	87.47	41.67
	OMT 3	84.58	21.91	87.72	24.11
	OMT 4	60.85	11.45	--	--
	OMT 5	94.98	4.21	--	--
	OMT 6	92.81	46.52	--	--
Family Medicine Preceptorship Ratings	FM 642	43.61	33.81	42.59	37.56
	FM 652	43.09	28.56	42.95	28.94
	FM 662	40.93	38.29	--	--
	FM 672	42.22	28.57	--	--
	FM 682	42.95	33.41	--	--
	FM 692	43.65	29.60	--	--

The relationship between the linear composites of course performance measures and traditionally used predictors of success in medical school can then be analyzed using multivariate multiple regression. In the review of literature concerning the prediction of medical school performance, the following consistent predictors of performance were identified: premedical GPA, MCAT Science and Quantitative scores, unweighted average of MCAT scores, quality of undergraduate institution, the student's age, and the extent of the student's previous employment. Other variables on which data were collected which may provide relatively independent information about medical school performance are measure of verbal or general aptitude and achievement (e.g., MCAT Verbal and General scores, English GPA) and achievement in the behavioral sciences (Behavioral Science GPA).

The analysis strategy will be to enter the traditionally used predictors into the regression first and the other predictor variables second. The dependent measures (the linear composites of course performance measures) will be stepped into the regression in the order in which they were listed in the performance structures formulated above. Given previous findings reviewed in Chapter II and given the sizeable variances of the predictor variables, it can be hypothesized that a multivariate relationship exists between the group of predictors and the group of composite course performance measures.

The statistical assumptions of multivariate multiple regression are (Finn, 1974):

1. All of the predictor and criterion variables are linearly related.

2. The residuals on the dependent measures are independent and follow a multivariate normal distribution with expected values of zero and a variance-covariance matrix Σ.

In order to test the assumption of linearity, bivariate scatterplots of selected predictor variables and course performance measures were done. In all of the plots which were done, the pairs of variables were linearly related. These scatterplots simply confirmed the well known observation that most aptitude and achievement measures tend to be related in a predictably linear fashion (e.g., Dawes & Corrigan, 1974).

Summary

Data on student performance in basic science, clinical medicine, and clinical skills courses were collected from students matriculating in the Classes of 1974 and 1975 at MSU-COM. Depending upon the course, course performance measures consisted of composites of scores on objective (i.e., multiple-choice and true and false) midterm and final exams, evaluations of clinical skills and performance, and in some courses, ratings of laboratory skills. Data on medical school selection variables such as MCAT scores and GPA's were also collected.

From analyses of course content, four performance structures were developed. In the three-factor structure (illustrated in Figure 3.4), course performance measures were hypothesized to load on the following factors:

1. Factor I: Acquisition of basic facts and skills,
2. Factor II: Acquisition of clinical principles, and
3. Factor III: Application of clinical principles.

This proposed structure resembles the learning and performance hierarchies proposed by Bloom and his colleagues (1956) and Gagne' (1974).

An alternative structure which directly tests the relationship of clinical performance to performance in other facets of the curriculum is the following:

1. Basic science knowledge,
2. Clinical principles,
3. Clinical applications, and
4. Clinical skills.

In this structure all clinical skills courses, no matter when they were given, were hypothesized to load on the single, clinical skills factor.

Two other structures were also proposed. They were: (a) a two-factor structure in which basic science courses were hypothesized to load on one factor, and the systems biology and clinical skills courses on the second factor; and (b) a single, general factor structure as proposed by Maatsch et al. (1978).

When the researcher knows *a priori* what structures are to be investigated, confirmatory factor analysis permits the analyst to specify on which factors the measures load and then to test the fit of this hypothesized structure to a set of data (Jöreskog & Lawley, 1968). Using the COFAMM program for confirmatory factor analysis (Sörbom & Jöreskog, 1976), the statistical models underlying these hypothesized performance structures can be estimated. The COFAMM program provides maximum likelihood estimates of the factor loadings, the correlations among the factors, and the unique variances of the measures. These estimates of the correlations among the factors can be used to answer the principal research questions of the study concerning the relationship between clinical and basic science performance.

The fit of the four models can be compared by doing sequential chi-square tests for goodness of fit (Jöreskog, 1974). Once the most theoretically and statistically appropriate model has been chosen, the multivariate relationship between the performance factors and typically used medical school selection variables can be estimated by using multivariate multiple regression (e.g., Finn, 1974).

In this study both the initial estimates of the factor analytic and multivariate regression models will be done on the data from the Class of 1974. The data from the Class of 1975 will be used only to cross-validate the relationships among the factors in the 1974 model (i.e., the phi matrix from the confirmatory factor analysis) and the multivariate regression equations estimated on the 1974 sample.

CHAPTER IV

RESULTS

In this chapter, the results of the data analyses will be reported and discussed. First, the bivariate relationships among the course performance measures which furnish the basis for the confirmatory factor analyses will be described. Second, the results of the confirmatory factor analyses and the estimated relationships among the hypothesized performance factors will be reported. Finally, the results of the multivariate regression analyses and the estimated relationships among the admissions predictor variables and the medical school performance factors will be reported.

Interrelationships Among the Course Performance Measures

It is well known that the observed correlation between two variables is dependent upon the reliabilities of the variables. That is, the maximum observed correlation between X and Y will be less than or equal to the square root of the product of their reliabilties, viz,

$$r_{xy} \leq \sqrt{r_{xx}r_{yy}} \qquad (4.1)$$

Thus, a low observed correlation between two failable measures may actually approach or even equal the highest possible correlation between them. As such, this restriction on the size of the observed correlations should be kept in mind when evaluating the magnitudes of the observed relationships among the course performance measures.

The correlations among the course performance measures are displayed in Tables 4.1 and 4.2. The course performance measures have been grouped according to the factor patterns hypothesized in Chapter III. Several aspects of the correlation matrix for the Class of 1974 (Table 4.1) are worthy of note: The largest observed correlations in the matrix are found among the basic science courses. The next highest relationships occur among the systems biology courses, which are also relatively highly correlated with the basic science courses. Performance in the clinical skills courses (Physical Examination, Clinical Science, and Osteopathic Manipulative Therapy (OMT) is more highly correlated with performance in the basic science courses than with performance in either (a) the systems biology courses or (b) other clinical skills courses. The most likely explanations of these findings is that the course performance measures for the basic science courses were composites of scores on objective tests while the clinical course performance

Table 4.1

Correlations Among Course Performance Measures--1974 Sample
(Decimal Points Omitted)

	PSL	HIS	ANT	BCH	PHM	MPH	PTH	CL PHM	HEM	NER	CAR	RES	URN	GI	GD	ORT	PSY	PHY EX	PHY DX	CLS 6	CLS 7	OMT 1	OMT 2	OMT 3	OMT 4	OMT 5	OMT 6	FM 642	FN 652	FM 662	FM 672	FM 682	FM 692
PHYSL	100																																
HISTO	64	100																															
ANT	71	74	100																														
BIOCH	73	71	72	100																													
PHARM	59	45	51	60	100																												
MICRO	65	65	65	72	68	100																											
PATH	67	57	55	64	62	68	100																										
CL.PHM	65	29	47	53	53	53	67	100																									
HEMATO	59	38	49	51	57	53	63	58	100																								
NEURO	73	62	66	71	64	65	72	63	60	100																							
CARDIO	61	40	48	49	51	51	64	70	63	74	100																						
RESPIR	68	57	57	64	56	67	71	65	60	78	73	100																					
URIN	61	46	55	53	53	56	54	60	53	75	64	70	100																				
GI	33	39	38	44	39	46	54	52	49	60	53	50	67	100																			
GROWTH	58	54	52	60	60	63	61	52	58	70	60	70	71	65	100																		
ORTHO	35	35	39	39	37	44	50	48	56	50	57	59	48	56	61	100																	
PSYCHO	54	51	54	58	55	68	60	47	64	70	51	60	60	60	68	46	100																
PHYEX	46	54	43	43	35	50	53	38	37	53	39	54	39	38	47	44	49	100															
PHYDX	46	52	55	43	42	47	55	24	46	46	38	41	26	26	31	28	40	34	100														
CLSC16	48	44	43	47	44	51	59	37	40	58	51	52	49	46	52	37	46	44	35	100													
CLSC17	25	22	28	20	32	25	25	30	43	42	30	42	31	33	35	36	47	25	13	08	100												
OMT 1	42	51	43	46	38	59	54	40	45	48	37	48	35	40	44	39	47	53	47	33	24	100											
OMT 2	38	49	53	41	41	41	41	20	20	35	40	43	27	14	35	30	34	19	47	26	34	29	100										
OMT 3	55	45	55	43	50	58	39	29	36	49	38	49	45	24	42	34	42	39	50	34	18	30	40	100									
OMT 4	31	38	38	26	26	35	38	25	43	34	28	28	29	34	37	28	42	27	23	10	33	39	27	16	100								
OMT 5	26	14	18	19	32	10	23	23	16	32	30	30	29	21	23	17	10	30	09	16	05	03	07	23	09	100							
OMT 6	15	19	24	14	27	28	25	20	26	23	18	26	28	27	29	29	22	27	25	23	21	31	21	40	10	21	100						
FM 642	09	13	22	04	09	16	-05	03	-07	-01	06	02	18	03	11	-19	08	03	04	-14	-03	07	20	05	18	-01	04	100					
FM 652	00	03	-01	09	13	13	06	08	08	19	18	21	12	11	11	05	08	06	10	08	08	01	-02	08	-02	00	17	12	100				
FM 662	23	30	22	12	24	26	20	21	11	23	33	35	18	08	21	17	09	31	25	14	17	28	36	15	17	18	22	24	22	100			
FM 672	-02	-09	03	03	04	09	03	14	17	19	18	17	08	17	10	20	21	13	12	12	26	10	01	14	02	-25	-02	09	14	09	100		
FM 682	13	15	12	16	16	21	24	15	29	22	25	23	09	15	18	19	13	26	19	20	19	20	31	08	13	02	04	15	23	29	28	100	
FM 692	37	41	43	36	29	35	40	20	29	32	15	40	39	29	29	24	31	29	39	36	22	21	32	36	26	17	16	06	10	22	18	16	100

Table 4.2

Correlations Among Course Performance Measures--1975 Sample
(Decimal Points Omitted)

	PSL	HIS	ANT	BCH	CL PHM	MPH	PTH	PHM	HEM	INT	END	NER	PHY EX	PHY DX	OMT 1	OMT 2	OMT 3	FM 642	FM 652
PHYSL	100																		
HISTO	61	100																	
ANAT	62	49	100																
BIOCHM	72	51	57	100															
PHARM	57	35	36	39	100														
MICRO	52	37	36	41	62	100													
PATH	64	33	37	52	65	63	100												
CLPHM	52	34	28	29	56	53	58	100											
HEMATO	35	27	25	31	44	38	52	56	100										
INTEG	43	34	27	36	54	53	59	57	40	100									
ENDOC	56	50	30	32	47	52	47	58	43	53	100								
NEURO	58	42	32	36	59	61	61	66	53	58	69	100							
PHYEX	57	51	51	48	39	46	49	43	32	43	37	41	100						
PHYDX	42	27	41	33	45	39	42	42	28	41	29	45	47	100					
OMT 1	36	33	45	39	39	39	35	32	41	31	20	41	50	42	100				
OMT 2	18	23	29	01	25	07	00	23	20	31	37	19	17	16	17	100			
OMT 3	31	25	51	18	11	22	11	30	24	18	23	29	40	36	32	22	100		
FM 642	00	02	07	-06	-03	-15	-04	-03	00	-15	-09	06	01	-09	11	03	09	100	
FM 652	17	07	20	08	19	08	03	-01	01	11	11	17	-01	09	03	03	17	08	100

measures were composites of mainly ratings of clinical skills, which generally have lower reliability than the objective test scores.

The correlation of .08 between Clinical Science 6 and Clinical Science 7 was an anomaly as both course performance measures were supposed to assess the performance of clinical skills in hospital settings. Clearly these variables are not measuring the same performance, and because of this, both were dropped from further analysis.

The correlations among the volunteer physicians' ratings of students' clinical performance which make up the course performance measures for the Family Medicine preceptorships were disappointingly low, reflecting variance and low rater reliability of these ratings. With the exception of the ratings for the last preceptorship (FM 692), these measures also fail to correlate with any of the other course performance measures in the matrix. The same was found for the course performance measures for OMT 5 to OMT 8. As a result, the measures for FM 642 to FM 682, and for OMT 5 to OMT 8 were not included in further analyses. The elimination of these courses should not unduly affect the results of the analysis as the performance of clinical skills should be adequately sampled by course performance measures from the physical diagnosis and examination courses, the four remaining OMT courses, and FM 692.

Results of the Confirmatory Factor Analyses

As discussed in Chapter III, a necessary but not a sufficient condition for properly identifying a structural model is to fix at least k^2 parameters in the Λ and Φ matrices of the factor analytic model. In order to do this, the following procedure suggested by Jöreskog and Lawley (1969) was used: The diagonal elements in the Φ matrix were fixed at unities and the loading of variables which were *not* hypothesized to load on a given factor were fixed at zeros.

The first model which was estimated was the four-factor performance structure in which the clinical skills courses were hypothesized to load on the same factor regardless of when they were offered (the model illustrated in Figure 3.4). The maximum likelihood estimates of the parameters in the factor pattern (Λ) matrix and their standard errors (in parentheses) are displayed in Table 4.3. The loadings on the first or basic science factor are both high and, within approximately plus or minus one standard error, are the same. The loadings on the second and third factors are likewise generally consistent, and all loadings are within plus or minus two standard errors of each other. With the exception of the course performance measures for OMT 4 and the Preceptorship course, the loadings on the fourth factor are relatively consistent but not as high

Table 4.3

Parameter Estimates and Standard Errors for Four-Factor Model
(Decimal points omitted)

Course		Factor			
		Basic science	Clinical principles	Clinical application	Clinical skills
Physiology		833 (09)	--	--	--
Histology		778 (09)	--	--	--
Anatomy		810 (09)	--	--	--
Biochemistry		833 (09)	--	--	--
Pharmacology		720 (10)	--	--	--
Microbiology		838 (09)	--	--	--
Pathology		807 (09)	---	--	--
Clinical Pharmacology		--	741 (10)	--	--
Hematopoetic		--	715 (10)	--	--
Neurology		--	904 (09)	--	--
Cardiovascular		--	810 (09)	--	--
Respiratory		--	868 (09)	--	--
Urinary		--	813 (09)	--	--
Gastrointestinal		--	--	745 (10)	--
Growth & Development		--	--	864 (09)	--
Orthopedics		--	--	684 (10)	--
Psychopathology		--	--	789 (09)	--
Physical Examination		--	--	--	613 (10)
Physical Diagnosis		--	--	--	662 (10)
Osteopathic	OMT 1	--	--	--	643 (10)
Diagnosis &	OMT 2	--	--	--	559 (11)
Manipulative	OMT 3	--	--	--	638 (10)
Therapy (OMT)	OMT 4	--	--	--	444 (10)
Family Medicine Preceptorship		--	--	--	499 (11)

as the loadings on the other three factors. These similarities in the magnitudes of the factor loadings lend credence to the idea of forming unit-weighted composites of the course performance measures in order to construct estimates of performance on each of the four factors for the regression phase of the analysis.

The estimated correlations among the *true scores* on factors are given in Table 4.4. These correlations reveal a number of interesting and significant findings. First, as indicated by the size of their standard errors, all were highly statistically significant. Second, the correlations were all high (i.e., .77 to .93) indicating a remarkable consistency among the facets of performance measured by these factors. These high correlations are similar to those reported by Maatsch et al. (1979) in their analysis of the performance of emergency physicians but greater than the "observed score" cannonical correlations reported by Sirotkin and Whitten (1978) in their study of medical students' performance. Third, the highest, and most surprising, correlation was between the basic science factor and the clinical skills factor ($\hat{\Phi}_{14}$ = .934) which shows a high degree of relationship between measures of knowledge in basic science and measures of clinical skill when they have in effect been "corrected for attenuation." This surprising finding supports the rejection of the null hypothesis of little or no relationship between clinical

Table 4.4

Estimated Correlations Among Factors and Standard Errors
(Decimal points omitted)

	Factor			
Factor	1	2	3	4
1 Basic science	1000			
2 Clinical principles	884 (03)	1000		
3 Clinical application	881 (05)	912 (03)	1000	
4 Clinical skills	934 (04)	790 (06)	769 (07)	1000

and basic science performance and answers the principle research question posed in Chapter I.

The likelihood ratio chi-square associated with the model was 381.25 with 246 degrees of freedom. The Tucker-Lewis coefficient for the model was .970 which indicates a very reasonable fit of the model to the data. Another indication of the good fit of the model is that all of the factor loadings were statistically significant (Jöreskog, 1971).

The three-factor model (schematically illustrated in Figure 3.4) proposed the following performance structure:

1. Basic knowledge and psychomotor diagnostic skills factor;

2. Clinical principles factor; and

3. Clinical application factor.

Interestingly, this conception receives empirical support from the previous model in that the extremely high correlation between the basic science and clinical skills factors indicates that some of the course performance measures which loaded on the clinical skills factor might just as well have loaded on the basic science factor of the four-factor model.

The estimated factor loadings and their standard errors for the three-factor model are shown in Table 4.5. As was the case for the four-factor model, all of the factor loadings were highly significant. Neither the reduction in the number of factors nor the redistribution of the clinical skills courses among the three remaining factors in the model appreciably affected the magnitudes of the factor loadings for courses loading on the same factors as they did in the previous model.

The estimated true score correlations among the factors are displayed in Table 4.6. As in the four-factor model, they were all large (i.e., .83 to .92) and highly statistically significant.

The likelihood ratio chi-square for the three factor model was 404.69 with 249 degrees of freedom. The Tucker-Lewis coefficient for the model was .966 which indicates almost as good a fit to the data as did the four factor model. As discussed in Chapter III, the increased degree

Table 4.5

Parameter Estimates and Standard Errors for Three-Factor Model
(Decimal points omitted)

Course		Factor		
		Basic knowledge	Clinical principles	Clinical application
Physiology		826 (09)	--	--
Histology		781 (09)	--	--
Anatomy		805 (09)	--	--
Biochemistry		837 (09)	--	--
Pharmacology		718 (10)	--	--
Microbiology		839 (09)	--	--
Pathology		810 (09)	--	--
Clinical Pharmacology		--	730 (10)	--
Hematopoetic		--	720 (10)	--
Neurology		--	902 (09)	--
Cardiovascular		--	802 (09)	--
Respiratory		--	868 (09)	--
Urinary		--	810 (09)	--
Gastrointestinal		--	--	740 (10)
Growth & Development		--	--	861 (09)
Orthopedics		--	--	678 (10)
Phychopathology		--	--	791 (09)
Physical Examination		593 (10)	--	--
Physical Diagnosis		601 (10)	--	--
Osteopathic	OMT 1	613 (10)	--	--
Diagnosis and	OMT 2	524 (10)	--	--
Manipulative	OMT 3	--	549 (10)	--
Therapy (OMT)	OMT 4	--	404 (11)	--
Family Medicine Preceptorship		--	--	395 (11)

Table 4.6

Estimated Phi Matrix and Standard Errors for Three-Factor Model (Decimal Points Omitted)

	Factor		
Factor	1	2	3
1 Basic Science	1,000		
2 Clinical Principles	880 (03)	1,000	
3 Clinical Skills	934 (04	799 (06)	1,000

of fit of a more complex model over a simpler model can be statistically evaluated by testing the difference between the chi-squares for both models referenced to the difference in their degrees of freedom. The difference in the chi-squares for the three- and four-factor models was a chi-square of 23.44 with three degrees of freedom. This value was significant at less than the .005 level which indicates that the four-factor model provides a statistically better fit to the data.

While the logic of the sequential chi-square test would suggest acceptance of the four-factor model over the three-factor model, the Tucker-Lewis coefficient for the models were within .034 of each other. Thus, in spite of the statistical significance of the difference between the chi-squares for the two models, the four-factor model with its separate clinical skills factor explained only 3.4%

more covariation among the course performance measures than did the three-factor model. However, since the relationship between performance courses with explicitly clinical content and performance in basic science courses is of primary interest, the four-factor structure was selected as the most appropriate model of the observed course performance data.

As discussed above, the results of the sequential chi-square test would "tell" the investigator to stop with the estimation of the three-factor model. However, for the sake of closure, a two- and a one-factor model were estimated. The high true score correlation between the basic science and clinical skills factors (i.e., $\Phi_{14} = .929$) and the clinical principles and clinical applications factors ($\Phi_{23} = .912$) suggested that the best fitting two-factor model would probably consist of (a) a basic knowledge and psychomotor diagnostic skills factor, and (b) a clinical principles and applications factor.

The parameter estimates and standard errors for the two- and one-factor models are shown in Table 4.7. As in previous models, the estimated factor loadings were both high and statistically significant. The high estimated correlation between the factors in the two-factor model ($\Phi_{12} = .889$) indicates, *at least in a statistical sense*, the dependence between performance on measures

Table 4.7

Parameter Estimates and Standard Errors for Two- and One-Factor Models (Decimal Points Omitted)

Course		Factor 1	Factor 2	Factor 1
Physiology		823 (09)		806 (09)
Histology		782 (09)		713 (10)
Anatomy		805 (09)		755 (09)
Biochemistry		837 (09)		791 (09)
Pharmacology		718 (10)		720 (10)
Microbiology		841 (09)		809 (09)
Pathology		810 (09)		821 (09)
Clinical Pharm			721 (10)	706 (10)
Hematopoetic			727 (10)	720 (10)
Neurology			894 (09)	885 (09)
Cardiovascular			791 (09)	752 (09)
Respiratory			862 (09)	849 (09)
Urinary			813 (09)	766 (09)
Gastrointestinal			689 (10)	639 (09)
Growth & Devel.			813 (09)	790 (09)
Orthopedics			647 (10)	609 (10)
Psychopathology			762 (09)	760 (09)
Physical Examination		594 (10)		598 (10)
Physical Diagnosis		601 (10)		554 (10)
Osteopathic	OMT 1	615 (10)		595 (10)
Diagnosis &	OMT 2	525 (10)		486 (11)
Manipulative	OMT 3		546 (10)	582 (10)
Therapy (OMT)	OMT 4		415 (11)	427 (11)
FM Preceptorship			407 (11)	437 (11)

of basic knowledge and clinical skills acquired during the first part of the curriculum and measures of the acquisition of clinical principles and the application of these principles later in the curriculum. The chi-square associated with the two-factor model was 415.60 with 251 degrees of freedom. The Tucker-Lewis coefficient for the model was .964 which indicates a very respectable fit.

The parameter estimates and standard errors for the single factor model are also shown in Table 4.7. The chi-square for this model was 468.73 with 252 degrees of freedom. The Tucker-Lewis coefficient was .953 indicating a remarkably good fit for a model which hypothesizes only a single factor to account for the covariation in the performance measures. Both the high degree of fit of the one-factor model and the high estimated correlations among the factors in the four-, three-, and two-factor models lend support to the Maatsch et al. (1978) conclusion that a single general ability factor underlies performance on both cognitive tests of medical knowledge and clinical simulations of diagnosis and case management. These findings also support Ebel's (1969) common sense criticism of learning and performance hierarchies, viz., that a student who knows more or, in Ebel's words, has a more well developed "structure of knowledge" will perform better at <u>all</u> levels of the hypothesized taxonomy or hierarchy.

In order for the results of the study to have external validity, the relationship among the factors (i.e., the estimated Φ matrix of the factor analytic model) should be approximately the same from sample to sample. While very little research has been done on what "factors" influence the magnitudes of these correlations, considerable work has been done on the variables which affect the magnitudes of the estimated factor loadings in the factor pattern matrix. Since the estimated correlations among the factors are theoretically equivalent to correlations among linear composites of the original variables weighted by a transformation of the factor loadings, it can be assumed that factors which affect the magnitudes of the factor loadings also affect the magnitudes of the correlations among the factors. According to Gorsuch (1974), some of these "factors" are:

1. The variances of the variables being analyzed;
2. The reliabilities of the variables; and
3. The estimated variance of the factor scores (i.e., the diagonal elements of the Φ matrix).

In the best of all possible worlds, the cross-validation of the Φ matrix would be carried out on another sample which contained measures on exactly the same variables which were included in the first sample. However, if data are not available on all of the variables which were factor analyzed in the first sample, or if additional variables are included in the model for the second sample, both of these conditions will likely cause some discrepancies in the estimates of the factor loadings of variables common to both samples. Thus, if the variables in the two samples are not equivalent for any of the above reasons, it will be more difficult to cross-validate the original model. In a sense, however, these differences are advantageous because their existence makes for a more stringent test of the "robustness" of the original model.

The model which was estimated on the data from the Class of 1975 differed in some respects from the model for the 1974 sample. First, attempts to collect data on courses which were missing for the Class of 1975 were not successful (partly due to the fact that coordinators for some of the courses have left the University). Consequently, the model for the Class of 1975 lacks the course performance measures which were used to construct the clinical application factor which was included in the 1974 model. However, since the estimated true score correlation between the clinical application and the clinical principles factor was almost unity (i.e., $\hat{\Phi} = .912$), these two factors could have been conceivably combined in the 1974 model resulting in a minimal loss of independent information. Second, data on two additional systems biology courses, Integumentary and Endocrine, were available for the 1975 sample but not for the 1974 sample.

As can be seen from the parameter estimates in Table 4.8, the estimated factor loadings are generally from .05 to .10 lower in magnitude than the comparable loadings in the factor pattern matrix for the 1974 sample (shown in Table 4.3). The estimated Φ matrix for the 1975 sample is displayed in Table 4.9. For purpose of comparison, the upper triangle of the matrix contains the comparable correlations from the 1974 four-factor model. While all of the estimated correlations for the 1975 sample are substantial and highly significant, they are all smaller than comparable correlations for the 1974 sample. However, the size of these estimated correlations (i.e., .707 to .822) and their high statistical significance support the conclusion of a high degree of relationship among the three hypothesized factors of medical school performance.

The chi-square associated with the model for the 1975 sample was 234.28 with 116 degrees of freedom. The Tucker-Lewis coefficient for the model was .929 which, while lower than that for the 1974 three-factor model, indicates that the model accounted for a substantial amount (i.e., 93%) of the covariation in the measures.

Table 4.8

Parameter Estimates and Standard Errors for Three-Factor Model--1975 Sample[a]

Course		Factor 1 Basic Science	Factor 2 Clinical Principles	Factor 3 Clinical Skills
Physiology		857 (08)		
Histology		623 (10)		
Anatomy		637 (10)		
Biochemistry		700 (09)		
Pharmacology		719 (09)		
Microbiology		704 (09)		
Pathology		768 (09)		
Clinical Pharm.			785 (09)	
Endocrine			763 (09)	
Integumentary			698 (09)	
Hematopoetic			621 (10)	
Neurology			857 (09)	
Physical Examination				781 (10)
Physical Diagnosis				646 (10)
Osteopathic Diagnosis & Manip. Therapy	OMT 1			632 (10)
	OMT 2			266 (11)
	OMT 3			490 (11)

[a]Decimal points omitted.

Table 4.9

Estimated Phi Matrix and Standard Errors for Three-Factor Model--1975 Sample[a] (Decimal Points Omitted)

	Factor		
Factor	1	2	3
1 Basic Science	1,000	884 (03)	934 (04)
2 Clinical Principles	822 (05)	1,000	790 (06)
3 Clinical Skills	815 (06)	707 (08)	1,000

[a]Comparable correlations for the Class of 1974 are displayed above the diagonal.

Results of the Multivariate Regression Analysis

In the review of the literature in Chapter II, the following variables were identified as the most consistent predictors of medical school performance: premedical GPA, premedical science GPA, MCAT Science and Quantitative scores, and the quality of the applicant's undergraduate institution. With the exception of ratings of the quality of the MSU-COM applicant's undergraduate school (which were not done because of the fact that some applicants had attended four or five different undergraduate schools before receiving baccalaureate degrees), data on all of these variables were collected on the students in this study. As shown in Table 3.4, additional data were collected on students' GPA's in different subject matter areas (e.g., biological science, English, behavioral science).

In order to provide the maximum amount of information, the set of predictor variables in a regression should be highly correlated with the criterion or dependent variable(s), and yet have the smallest possible correlations among themselves (Kerlinger & Pedhazur, 1973). The following set of predictor variables should meet this desideratum yet still include some of the typically used predictor variables: (a) measures of scientific competence (Science GPA, MCAT Science, Biology GPA), (b) measures of verbal competence (MCAT Verbal and English GPA), (c) a measure of achievement in the behavioral sciences (Behavioral Science GPA).

In order to test the multivariate relationship between the predictor variables, on the one hand, and the composite performance scores, on the other, Finn's MULTIVARIANCE program for multivariate analysis of variance was used. In this analysis, the predictor variables were entered into the multivariate regression in the approximate order of their "popularity" as medical school selection variables (cf., Chapter II), i.e., (a) Science GPA, (b) MCAT Science, (c) Biology GPA, and (d) the indicators of non-biological and physical science achievement and aptitude (i.e., MCAT Verbal, English GPA, and Behavioral Science GPA). As suggested by Bock (1968), the composite performance measures were entered into the regression in a temporal order (i.e., in the order in which the groups of courses were offered in the MSU-COM curriculum: (a) Basic Science, (b) Clinical Principles, (c) Clinical Skills, and (d) Clinical Application).

As was proposed in Chapter III, the composite measures of medical school performance (i.e., estimates of factor scores on the four performance factors) were formed by computing unit-weighted linear composites of the T-scores on the course performance measures. For example, the composite score on the clinical application factor was calculated by adding together the T-scores on the course performance measures for the Gastrointestinal, Growth and Development, Orthopedics, and Psychopathology courses. The means, standard deviations, and ranges of the predictor variables and the composite performance scores are shown in Table 4.10. As can be readily seen from the standard deviations and ranges, all of the variables display considerable variability.

The multivariate F-ratio for the null hypothesis of no overall association between the predictor and criterion variables was 5.19 (d.f. = 24 and 259; $p < .0001$), indicating a highly significant multivariate association. The contributions of each of the predictor variables to this

Table 4.10

Summary Statistics for Predictor and Criterion Variables--
1974 Sample

Variable	Mean	Standard deviation	Range ($\bar{X}$ ± 2 S.D.)
Basic Science Factor	350.00	58.27	233-467
Clinical Principles Factor	300.00	50.70	199-401
Clinical Skills Factor	350.00	45.84	258-442
Clinical Application Factor	200.00	33.32	133-267
Science GPA	2.97	0.43	2.11-3.83
MCAT Science	511.71	95.75	320-703
Biology GPA	3.23	0.48	2.27-4.00
MCAT Verbal	494.20	84.75	325-664
English GPA	2.90	0.50	1.90-3.90
Behavioral Science GPA	3.25	0.63	1.99-4.00

association can be assessed by examining the "stepwise" F-ratios as each predictor variable is added to the multivariate regression model. The logic behind this sequential testing procedure is to eliminate predictors which do not provide significant additional information beyond that provided by predictors which are already in the regression equation. The multivariate F-ratios for these tests are displayed in Table 4.11. From the results of these tests, it can be clearly seen that the measures of verbal aptitude and behavioral science achievement do not contribute significant additional information to the

Table 4.11

Stepwise Tests of Contributions of Predictor Variables--
1974 Sample

Predictor	F	d.f.	Significance
Science GPA	10.06	4, 79	.0001
Science MCAT	13.09	4, 78	.0001
Biology GPA	7.60	4, 77	.0001
MCAT Verbal	0.66	4, 76	.6188
English GPA	0.61	4, 75	.6593
Behavioral Science GPA	0.84	4, 74	.5037

prediction of the criterion variables beyond that already provided by the three measures of science aptitude and achievement which were entered earlier. One possible reason for this finding is that the English and Behavioral Science GPA's were calculated on the basis of fewer courses in these subjects than the Science and Biology GPA's and, hence, probably had lower reliabilities. As a result of their lack of additional significance, the model was reestimated deleting these predictor variables and all statistics reported from this point on are based upon this reestimated model.

The sequence of step-down F-ratios provided by the MULTIVARIANCE program enables the researcher to test the relationship between the q predictors and the first criterion variable, the predictors and the second criterion variable controlling for scores on the first criterion, the predictor and the third criterion controlling for scores on the first two criteria, and so forth (Finn, 1974). The logic behind this sequential testing procedure is that criterion variables which do not provide significant additional information beyond that provided by criterion measures which have already been entered into the regression are candidates for deletion from the model.

The step-down F-ratios are shown in Table 4.12. The F-ratio for the null hypothesis of no association between the predictor variables and the performance on the Clinical Application factor was 1.07 ($p < .3661$) which indicates that after controlling for performance on the other three factors there was a non-significant association between the predictor variable and performance on the clinical application factor. After controlling for

Table 4.12

Step-Down Tests of Association for Criterion Variables

Criterion	F[a]	Significance
Basic Science	46.25	.0001
Clinical Principles	3.26	.0258
Clinical Skills	2.58	.0594
Clinical Application	1.07	.3661

[a](D.f. = 3, 80).

performance on both the Basic Science and Clinical Principles factors, the contribution of the Clinical Skills factor was marginally significant ($F = 2.58$; $p < .059$). Due to the high degree of association among the criterion variables (especially between the Clinical Principles and Clinical Application factors), these results are not surprising.

The prediction of each of the performance factors individually can be evaluated by looking at the univariate regression statistics for each. The univariate multiple R's, R^2's, and their associated levels of significance are displayed in Table 4.13. The raw score regression coefficients, standard errors, squared zero-order correlations, and standardized regression coefficients (betas) are shown in Table 4.14. The multiple R^2's in Table 4.14 show that the linear combination of the three predictors which were retained in the model accounted for 63% of the variation in the Basic Science factor scores and over 40% of the variation in each of the two clinical factor scores.

The conventional explanation of these findings would be that test-taking ability, ability to memorize information, and test-wiseness in undergraduate school predict similar qualities in professional and graduate school. However, two aspects of the data themselves do not support this wary view of standardized tests and grade point

Table 4.13

Univariate Regression Statistics for Criterion Variables--
1974 Sample

Criterion	Multiple R	Multiple R^2	F[a]	Significance
Basic Science	.796	.634	46.25	.0001
Clinical Principles	.697	.486	25.22	.0001
Clinical Skills	.646	.417	19.05	.0001

[a](D.f. = 3, 80).

Table 4.14

Regression Coefficients and Standard Errors--1974 Sample

Factor score/ predictors	Raw score regression coefficient	Standard error	Standardized coefficient (Beta)
Basic Science:			
Science GPA	7.639	12.25	.057
MCAT Science	0.311[a]	0.04	.510
Biology GPA	55.862[a]	10.96	.460
Clinical Principles:			
Science GPA	34.976[a]	12.63	.298
MCAT Science	0.195[a]	0.04	.367
Biology GPA	26.682[a]	11.31	.253
Clinical Skills:			
Science GPA	-3.608	12.17	-.034
MCAT Science	0.133[a]	0.04	.277
Biology GPA	52.096[a]	10.89	.546

[a]($p < .05$).

averages. First, performance on the Clinical Skills factor was heavily weighted in favor of ratings of psychomotor skill performance as opposed to performance on written examinations (i.e., ratings of clinical skills were weighted 1.5 to twice as much as performance on written tests in determining the students' final grade in the OMT courses). Second, as can be seen from the correlations among the predictors and criterion variables in Table 4.15, there were surprisingly low correlations between Science GPA and Biology GPA on the one hand and MCAT Science on the other hand. These low correlations indicate that the predictor variables are measuring different (but not entirely independent) aptitudes or achievements.

The relationships among the predictors can be conceptualized in terms of the percentage of overlap among the variances of any two predictors. This percentage of overlap can be estimated by the square of the zero-order correlation between them (the so-called "coefficient of determination"). These squared zero-order correlations are shown above the diagonal in Table 4.15. Science GPA and Biology GPA obviously have a high degree of overlap as their relationship is a "part-whole" correlation (i.e., the Science GPA includes grades from biology courses). On the other hand, Science MCAT scores share relatively little variance with either of the two grade point averages.

Table 4.15

Correlations Among Predictor and Criterion Variables--1974 Sample[a]

Variable	Basic Sci.	Princ.	Skill	Appl.	Sci. GPA	MCAT Sci.	Bio. GPA	MCAT Verb.	Eng. GPA	B.S. GPA
Basic Science	1.000	.653	.643	--	.240	.399	.381	--	--	--
Principles	.808	1.000	.444	--	.311	.252	.286	--	--	--
Skills	.802	.663	1.000	--	.157	.157	.346	--	--	--
Application	.710	.817	.642	1.000	--	--	--	--	--	--
Science GPA	.490	.558	.396	.543	1.000	.065	.434	--	--	--
MCAT Science	.632	.502	.396	.442	.255	1.000	.055	--	--	--
Biology GPA	.617	.535	.588	.542	.659	.234	1.000	--	--	--
MCAT Verbal	.336	.185	.238	.213	.168	.474	.272	1.000	--	--
English GPA	.280	.301	.303	.249	.248	.149	.313	.077	1.000	--
Behav. Sci. GPA	.272	.247	.242	.191	.402	.315	.337	.277	.317	1.000

[a]Squared zero-order correlations for variables included in the final model are shown above the diagonal.

From the squared correlations in Table 4.15, it can be seen that: (a) MCAT Science accounted for an approximately equal amount of variation in all of the criteria; and (b) in spite of the part-whole correlation between them, Biology GPA accounted for substantially more variance than did Science GPA on the Clinical Skills factor. The magnitudes of the squared zero-order correlations are somewhat related to the magnitudes of the standardized regression coefficients shown in Tables 4.14 (although for reasons to be discussed below this is definitely not always the case in regression analyses with correlated independent variables).

The standardized regression coefficient gives the estimated change in the dependent variable in standard deviation units when the predictor variable is increased by one standard deviation *and* the values on all of the other predictors are held constant. In spite of their greater sensitivity to the sample variances of the predictors, the betas are easier to compare and interpret than raw score regression weights because they are expressed in terms of the same unit of measurement.
As the degree of correlation (or collinearity) among the predictors increases, the magnitudes of the beta weights become more sample dependent, and hence, in order to be taken seriously, should be cross-validated on another

sample. With these caveats in mind, the sizes of the beta weights can be interpreted as reflecting the amount of "independent" contribution a predictor makes to the dependent variable above and beyond the contributions made by the other predictors (i.e., when the other predictors are held constant). Using this interpretation, it can be seen that (a) the three independent variables contribute approximately equally to the prediction of performance on the Clinical Principles and (b) the Biology GPA is a reasonably good predictor of Clinical Skills performance.

The strength of the overall multivariate association between a set of predictor and a set of criterion variables can be estimated by the series of canonical correlations between the sets (Finn, 1974). According to Finn, the canonical correlation is the simple correlation of two random variables (say, w_1 and v_1) each of which is a linear combination of the predictor and criterion variables, respectively. The weights for these linear combinations are chosen so as to maximize the zero-order correlation between the linear combinations. The weights for the second canonical correlation are chosen so as to maximize the correlation between a second pair of linear combinations (say, w_2 and v_2) subject to the restrictions that $r_{w_1 w_2} = 0$ and $r_{v_1 v_2} = 0$ (i.e., that the linear combinations of each set are orthogonal). If the measures are assumed to be standardized to unit variance, the overall percentage of variance accounted for in the *p* criterion measures by the *q* predictors is,

$$P = \frac{100\ \Sigma\ R_i^2}{p} \tag{4.2}$$

where R_i^2 = the *i* th canonical correlation.

The number of canonical correlations which will be yielded by the analysis is the minimum of the number of independent or dependent variables. In this case, with three independent and three dependent variables, three canonical correlations will be generated. The significances of the correlations can be tested by sequential chi-square tests, testing the significance of the last correlation first, the last two correlations second, and so forth. As was the case with the step-down F tests discussed earlier, the logic of this test is to stop testing when a significant result is encountered. The estimates of the canonical correlations and the results of these sequential tests are shown in Table 4.16. As can be seen from these results, all three canonical R's are significant. Applying Equation 4.2 to the canonical R's, it can be calculated that 27.6% of the variance in the three criterion scores is accounted for by the predictor variables.

Table 4.16

Canonical Correlations and Significance Tests--1974 Sample

Canonical correlation	R_i	Significance test of R_i	χ^2	Significance
1	.799	1 through 3	96.77	.0001
2	.339	2 through 3	16.00	.0031
3	.275	3 through 3	6.27	.0123

Due to the varying degrees of dependence of the regression statistics on the sample variances of the measures, and on omnipresent sampling variability, it is important to cross-validate these statistics on a second sample when possible. Using the procedure described above, a multivariate regression analysis was performed on the data for the Class of 1975. As in the 1974 sample, the overall test of the null hypothesis of no association between the three predictor and three criterion measures was significant (F [9, 219] = 5.255, $p < .0001$). In contrast to the results of the 1974 analysis, the stepwise tests of the contributions of the predictor variables showed that Biology GPA did *not* contribute significantly to the prediction of the criterion variables above and beyond Science GPA and MCAT Science (F [3, 91] = 9.61, $p < .0001$) and Science GPA (F [3, 92] = 6.46, $p < .0001$) which were both highly statistically significant.

The results of the step-down tests are shown in Table 4.17. The last criterion variable which was entered into the regression equation was Clinical Skills assessment. The step-down F for this factor was 2.32 ($p < .08$) which indicates a marginal contribution to the regression equation above and beyond the Basic Science and Clinical Principles factors which were entered earlier. In contrast to the results of the 1974 analysis, the addition of the Clinical Principles factor was not significant. In spite of this, the logic of the step-down testing procedure would recommend retention of this measure in the model because the Clinical Skills factor (which was entered later) was retained.

Table 4.17

Step-Down Test of Association for Criterion Variables--
1975 Sample

Variable	F	Significance
Basic Science	13.839	.0001
Clinical Principles	0.877	.4564
Clinical Skills	2.325	.0802

The results of the univariate regressions are shown in Table 4.18 and the regression weights are displayed in Table 4.19. In spite of the insignificant contribution of Biology GPA to the regression, the model was not reestimated.

Table 4.18

Univariate Regression Statistics for Criterion Variables--
1975 Sample

Variable	Multiple R	Multiple R^2	F	Significance
Basic Science	.558	.311	13.84	.0001
Clinical Principles	.334	.111	3.85	.0122
Clinical Skills	.500	.250	10.20	.0001

The rationale behind this decision was to allow for a direct comparison between the univariate regression weights for both the 1974 and 1975 samples. (Because of the lack of the significant contribution of Biology GPA, the estimates of multiple R's will not differ a great deal from those displayed in Table 4.18.)

The multiple R^2's (i.e., the proportions of variance in the criterion variables accounted for by the linear combinations of the predictor variables) while highly statistically significant were substantially less than those from the 1974 sample. For example, the three predictor variables accounted for 63.4% of the variation in the Basic Science performance in the 1974 sample but only 31.1% (or about half as much) of the variation in the performance in Basic Science in the 1975 group. Possible explanations of these precipitous decrements in the R^2's can be found in the correlations among the predictor and

Table 4.19

Regression Coefficients and Standard Errors--1975 Sample

Factor score/ predictors	Raw score regression coefficient	Standard error	Standardized coefficient (Beta)
Basic Science:			
Science GPA	28.290	18.78	.209
MCAT Science	0.256[a]	0.05	.434
Biology GPA	6.119	17.29	.051
Clinical Principles:			
Science GPA	19.330	16.07	.189
MCAT Science	0.091	0.05	.205
Biology GPA	2.786	14.79	.031
Clinical Skills:			
Science GPA	29.107[a]	11.03	.382
MCAT Science	0.107[a]	0.03	.322
Biology GPA	-5.341	10.15	-.079

[a]($p < .05$).

criterion variables, and in the variances and reliabilities of the factor scores.

As can be seen from Table 4.20, the correlations among the predictor variables and the factor scores are considerably smaller for the 1975 sample than for the 1974 sample. This is likely due, in part, to the lower variances of the factor scores (shown in Table 4.21 along with their 1974 counterparts), and in part to the lower reliabilities of the factor scores for the 1975 sample.

Summary

This section contains a summary of the results of the analyses. In order to avoid repetition, the results are discussed within the context of earlier findings in the first part of the next chapter.

The strongest relationships among the course performance measures occurred between the measures for the basic science courses, on the one hand, and (a) the systems biology courses and (b) the clinical skills courses, on the other hand. The observed correlations among all of the course performance measures were generally moderate to high.

In order to estimate the degree of relationship among the different facets of performance in medical school which were discussed in Chapter III, confirmatory, maximum likelihood factor analyses were done on the course performance

Table 4.20

Correlations Among Predictor and Criterion Variables--
1975 Sample

Variable	Basic Science	Principles	Skills	Science GPA	MCAT Science	Biology GPA
Basic Skills	1.000					
Principles	.719	1.000				
Skills	.659	.560	1.000			
Science GPA	.353	.263	.398	1.000		
MCAT Science	.503	.262	.386	.243	1.000	
Biology GPA	.370	.252	.334	.780	.359	1.000

Table 4.21

Summary Statistics for Predictor and Criterion Variables--
1975 Sample

Variable	Mean	Standard deviation[a]	Range ($\bar{X}$ ± 2 S.D.)
Basic Science	350.00	53.15 (58)	244-456
Clinical Principles	250.00	40.04 (51)	210-330
Clinical Skills	200.00	29.90 (46)	170-230
Science GPA	3.09	0.39 (0.43)	2.31-3.87
MCAT Science	521.07	90.02 (96)	341-701
Biology GPA	3.24	0.44 (0.48)	2.36-4.00

measures for the Class of 1974. The four-factor structure which proposed Basic Science, Clinical Principles, Clinical Application, and Clinical Skills factors fit the observed relationships among the measures significantly better than the other models which posited fewer factors.

The estimated true score correlations among the factors were both high (i.e., .77 to .93) and statistically significant. Contrary to the hypothesis of little or no relationship between basic science course performance and the performance of clinical skills, the estimated correlation between the factors which measured these performances was .927. The correlations among the other factors were also high which strongly suggests that performance across both clinical and basic science areas in the curriculum is consistent.

In order to test the generalizability of these relationships, a three-factor structure was estimated on similar course performance measures from students in the entering class of 1975. As in the previous model, this hypothesized structure included Basic Science, Clinical Principles, and Clinical Skills factors. (The Clinical Applications factor was not included in the model due to the lack of data for most of these courses.) While slightly lower than those for the 1974 sample, the estimated true score correlations among the factors were also high (i.e., .71 to .82) and statistically significant. The Tucker-Lewis coefficient for the model was .93, which indicated an acceptable fit of the model to the data.

Three "clusters" of independent variables were chosen to be used to predict performance on the four criterion factors. These clusters were: measures of science aptitude and achievement (i.e., Science GPA, Science MCAT, and Biology GPA), verbal ability and achievement (i.e., MCAT Verbal and English GPA), and behavioral science achievement (Behavioral Science GPA). The results of a multivariate regression analysis showed that performance on the criterion measures was significantly related to the science predictors only. The results of the series of sequential, step-down F tests similarly demonstrated that the Clinical Applications factor did not contribute to the regression above and beyond the other three performance factors which were entered earlier. Hence, the verbal and behavioral science predictors and the Clinical Applications factor were deleted from the analysis and the regression model was reestimated.

The results of the univariate regression analyses showed that the science predictors accounted for 63% of the variance on the Basic Science performance factor; these predictors also accounted for 49% and 42% of the variation in the Clinical Principles and Clinical Skills factors, respectively. Together the predictors accounted for 28% of the variation in the three criterion measures.

While the associations between the same predictor and criterion variables were also highly statistically significant in the 1975 sample, the magnitudes of the associations were smaller. That is, the science predictors accounted for an estimated 12% of the variance in the criterion variables in the 1975 sample. Similarly, the univariate R^2's between the predictors and each criterion were also smaller than in the 1974 sample, i.e., .31 for Basic Science, .11 for Clinical Principles, and .25 for Clinical Skills. One logical explanation of the shrinkage of these statistics is the decreased variances and, hence, restrictions in range in the predictors.

CHAPTER V

CONCLUSIONS AND IMPLICATIONS

In this chapter the conclusions of the study and their implications for medical school admissions policy and further research will be enumerated. First, the results of the investigation will be briefly summarized and discussed within the context of previous research. Second, the conclusions and limitations of the study will be outlined and the potential generalizability of the results will be discussed. Third, suggestions for further research will be made. Fourth, the implications of the results for medical school admissions policy will be developed.

Discussion and Conclusions

The principal research question which was investigated by this study concerned the relationship among facets of performance in medical school. In Chapter III, four models of the structure of performance in osteopathic medical school were identified. These models were formulated on the basis of (a) course content; (b) the distribution of contact hours devoted to basic science, clinical science, and clinical skills instruction; and (c) the sequencing of these courses in the curriculum. The models, how they were formulated, and medical student performance data which were collected to test the hypothesized relationships are described in detail in Chapter III. The relationships among the facets of performance specified by the models were estimated by using maximum likelihood confirmatory factor analysis. This technique permits the researcher to specify on which latent factors the observed variables should load, and provides maximum likelihood estimates of the factor loadings, the "true score" correlations among the factors, and the unique variances of the measures.

The four-factor model, which included a separate clinical skills performance factor, fit the data best. The correlations among the four factors hypothesized by the model were all high (i.e., .77 to .93) and highly statistically significant. Using data from the entering class of 1975, these relationships were successfully cross-validated. The estimated correlations among the three hypothesized factors in the 1975 sample were also high (i.e., .71 to .82) and statistically significant.

These strong relationships among didactic and clinical skills' performance (which were measured by objective tests, and ratings of "hands-on" clinical performance) demonstrated a substantial degree of consistency of student performance across both medical school subject matter domains and measurement methods. These results are consistent with the moderate to high canonical correlations among performance in basic science and clinical courses reported by Sirotkin and Whitten (1978) for students in a similarly structured curriculum at another university. The results are also consistent with the strong relationships among scores on objective tests measuring clinically-relevant basic science knowledge and clinical knowledge, and performance on clinical simulations reported by Maastch et al. (1978, 1979) for practicing physicians. It should be stressed that the relationships in the studies described above and in the current study are correlational rather than causal relationships. Should an investigator wish to test for unidirectional links between different types of performance, structural equation or path models would be an appropriate alternative analytic strategy.

Unit weighted indices of performance on the four factors were developed by summing T-scores from the courses which loaded on each of the four factors. The results of the multivariate regression analyses showed that science GPA, MCAT Science, and Biology GPA were significantly related to scores on the four performance factors in both the 1974 and 1975 samples. The relationships in both samples were generally higher than those previously reported in similar studies from allopathic medical schools (see review in Chapter II). However, the strength of relationship between the predictor variables and performance factors was higher in the 1974 than in the 1975 sample.

The probable explanation of this difference in the findings was alluded to in Chapter II: It is well known (e.g., Gough, 1979) that schools of allopathic medicine place a great deal of emphasis on objective measures of achievement and aptitude in selecting applicants. Thus, the range of these predictor variables has been severely censored in allopathic student populations leading to attenuated relationships between these predictors and later medical school performance.

The principal conclusions of the study can be briefly summarized as follows:

1. Medical school performance is consistent across both subject matter domains and methods of measurement.

2. Consistent with the results of recently reported studies, basic science and clinical performance are more strongly related than had previously been reported.

3. Given a population with a wide variation in scores on objective predictors of medical school performance, these measures are substantially related to subsequent medical school performance.

Limitations of the Study and of the Generalizability of the Results

MSU-COM was the first college of osteopathic medicine to be established in the United States in over 50 years, and after its sister allopathic school at the University, the College of Human Medicine, the first new medical school to be established in Michigan in over 50 years. Because they were not as tied to tradition, both of these schools began their existences trying out new curricula, methods of teaching, and admissions criteria.

As discussed in Chapter I, the first few classes admitted to COM contained a relatively high percentage of hithertofore "non-traditional" medical students, i.e., higher percentages of minorities, women, applicants with predominantly non-science academic backgrounds, and older students preparing for a second career. These non-traditional students were admitted partly because the admissions committee at the time was not weighting GPA, MCAT scores, and other objective predictors of performance, as heavily as indicators of social commitment, previous health related activities, and motivation toward a career in osteopathic medicine. In 1975, however, the Committee began to place more weight on objective predictors of performance, and this emphasis (while still not as heavy as at most allopathic colleges) has continued to increase. Concomitant with this increasing emphasis, the premedical grade point average of new classes has risen, mean age has dropped, and fewer students with predominantly non-science academic backgrounds are being admitted.

The reason which was hypothesized for the strong findings yielded by the regression analyses in comparison to similar results from allopathic colleges was the greater range in values on the predictor variables. Hence, the limits to the generalizability of the results of the regression analyses are essentially related to the *range* of scores within which applicants are selected for admission to a particular school. If only applicants with high GPA's (e.g., 3.40 or better) and/or high MCAT scores (e.g., 1.5 S.D. above the mean or better) are admitted, these variables will not as reliably predict performance *within this range* of scores.

A second limitation of the study is the following: While the study included explicit and fairly reliable measures of clinical performance, all of these measures (with the exception of the ratings of students' preceptorship performance from the last Family Medicine preceptorship) were made during on-campus courses and clinical experiences. It certainly can be questioned whether this performance can be legitimately generalized in order to make inferences about clerkship performance, and more questionably, about post-graduate clinical performance.

This difficulty in findings measures of clinical performance which generalize across cases has been discussed by Elstein et al. (1978) who found physicians' clinical problem solving behavior to be disease or case specific. Thus, the low reliability of clinical performance measures across cases and across time may be due to the varied nature of medical practice.

Third, this study employed the old MCAT test as a predictor of student performance. Clearly, an important question is whether or not the same result would be obtained from the recently developed, new MCAT test. Until these validity studies are conducted, the generalization of the results of the current study to applicant populations who have been administered the new test should be done with caution.

Before the recent implementation of new rating forms, hospital-based physicians were asked to rate third-year students on their clinical performance and attitudes using "global" rating scales. These ten-point scales had very low variances and were highly negatively skewed. Hence, data from them were not included in this study. On the other hand, revised rating forms consist of short descriptions of the clinical skills, attitudes, and behaviors to be rated. This revised format asks the rater to identify the descriptor which best describes the student's level of skill, professional behavior, or attitude from among five or six alternatives. While the data from these revised scales have not yet been formally analyzed, it would seem *a priori* that this new rating format would increase the reliable variance of the ratings of clerkship performance and would probably be worth including in a future study. Another method for assessing clinical competence is through ratings of students' diagnostic and case management behavior with simulated patients. Recent research by Elstein et al. (1978), Maatsch et al. (1978, 1979), and Tinning (Note 3) has demonstrated the feasibility of this approach.

While unfortunately lacking measures of clinical clerkship performance, the study did adequately sample basic science and clinical performance (including osteopathic diagnostic and treatment skills) during the first two years of osteopathic medical school. To the extent that other osteopathic and allopathic medical curricula contain approximately the same "balance" of basic science and clinical instruction (this condition would be more applicable to osteopathic than allopathic schools because of the training in OMT clinical skills during the first two years), the results concerning the relationships among clinical and basic science performance should be generalizable to these schools. In addition, results should be particularly generalizable to schools which have organ systems curricula and/or substantial amounts of clinical input during on-campus training.

Most of the students in the entering class of 1974 are currently in family or general practice. While the specialty choices of students matriculating in later classes will not be known for another year or more, it would be both interesting and worthwhile to investigate the relationship between both premedical predictors and measures of medical school performance, on the one hand, and specialty choice when it is known, on the other. The "conventional wisdom" among the COM faculty is that students with higher premedical GPA's and MCAT scores will choose specialty practice over general practice. The variable weights given to these measures in admitting students during the past ten years has conveniently "set the stage" for a longitudinal test of this hypothesis.

Suggestions for Further Research

Based upon the limitations of the study discussed in the last section, the following suggestions for further research can be offered: Data on clinical clerkship performance from the revised rating instruments, and, if possible, ratings of students' clinical work with simulated patients should be included in a future study. The relationship between premedical predictors and medical school performance, on the one hand, and specialty choice, on the other, could also be profitably studied.

Using the conservative method of estimating the common variance between the set of predictor and the set of criterion variables described in Equation 4.2 (c.f., Stewart & Love, 1968), it was found that the predictor variables accounted for 28% of the variation in the criterion variables in the 1974 sample and only 11% of the variance in the performance variables in the 1975 sample. Consequently, it would be useful to investigate the predictive validity of non-academic predictors, such as study habit inventories, problem solving instruments, reading tests, and social-psychological instruments designed to measure the applicant's ability to relate to others (e.g., the Krupka et al. empathy scale, the Myers-Briggs). In this vein, MSU-COM in cooperation with OMERAD has begun the development of a problem solving skills assessment for COM applicants. Of specific use to colleges of osteopathic medicine would be an instrument designed to measure the applicant's interest in and commitment toward a career in osteopathically-oriented medical care in a family practice setting. This scale would, however, require considerable development and validation work so as to avoid "socially desirable" responses on the part of applicants who simply wish to get into a medical school regardless of its philosophy of medical practice.

An investigation of unidirectional links among performance in the areas investigated in this study would be an interesting adjunct to the present relational analyses which used confirmatory factor analysis. This investigation could be carried out by using the LISREL IV computer program developed by Jöreskog and Sörborn (1978).

Implications

The most important findings of this study were lack of support for the hypotheses of (a) little or no relationship between clinical and basic science performance and (b) little or no relationship between objective predictors and subsequent medical school performance in populations of students with wide ranges of scores on these measures. As mentioned above, these findings hold few if any implications for medical schools who select applicants from the upper ranges on distributions of applicant talent. The findings, however, hold wide-ranging implications for virtually all medical schools which are attempting to broaden the ethnic, socioeconomic, and student background compositions of their classes.

The explanation for this is not often discussed outside faculty and administrative conclaves but must be made explicit for a clear understanding of the issue: Prestigious medical schools, such as Harvard, Michigan, and the University of California Medical School (which was studied by Gough et al., 1964), virtually have their choice of unquestionably academically qualified minority applicants. On the other hand, other less prestigious schools either nationally or within the applicant's own state must draw from lower areas of the distributions of scores on objective measures in order to recruit more than a token number of minority applicants. This, in combination with the admission of the other types of "non-traditional" applicants discussed above, is why there is now a wider range on both predictor and criterion measures than existed 10 years ago (c.f. McGuire, 1977).

To say that degree of skin pigmentation "causes" lower academic performance is to misspecify the casual model. However, variables which may indeed contribute to lower academic performance are clearly confounded with race. That is, poorer academic preparation in high school and college, poorer academic self-concept, poorer study habits, and a number of other variables have been shown to be related to both race and later academic performance.

The recent Bakke decision of the U.S. Supreme Court has allowed medical and other professional schools to explicitly consider race as an admissions criterion. In terms of the results of this study, what recommendations can be made for doing this? As discussed in Chapters I and II, medical schools typically have a two-screen admissions procedure. The first screen is usually based upon some combination of objective predictors of medical school performance. Applicants whose combined scores exceed the school's cutoff point pass the first screen and are invited for a personal interview. Based upon privileged communications with admissions officers and faculty of other medical schools, the author learned that some medical schools consider race as part of the "first screen" by assigning "admissions points" to minority applicants in such a way as to compensate for lower scores on the objective predictors. The results of this study imply, however, that in doing this, a school which admits a wide range of applicants may be admitting minority students whose GPA's and MCAT scores have been "over compensated for" and, hence, students whose scores are predictive or low performance and even failure in medical school. In admitting an applicant with low enough scores so as to predict probable low performance or failure (the applicable cutting scores would have to be empirically determined for each school individually), the school is neither doing itself nor the student a favor.

Perhaps a better and fairer way (for all applicants) of explicitly considering ethnicity is to lower the cutting scores on the first screen for *all* applicants and to assign the "ethnicity points" to minority applicants *after* they have passed the first screen.

The results of this study also suggest an additional revision of the admissions procedure which would benefit all applicants regardless of ethnic background. The results of the regression analyses in this study showed that while a representative set of typically used predictor variables accounted for substantially higher percentages of variance in criterion measures of performance than had previously been reported, the predictors by no means accounted for most of the variance in the criterion measures. This result implies that other predictors (such as those discussed above), while correlated with the typically used predictors, demonstrate a potential for offering additional independent information about criterion performance.

REFERENCE NOTES

REFERENCE NOTES

1. Hunter, J. E., & Gerbing, D. W. *Unidimensional measurement and confirmatory factor analysis.* Paper No. 20. East Lansing, Mich.: Institute for Research on Teaching, Michigan State University, 1979.

2. Tinning, F. C. *An experimental study investigating the effects of real and simulated clinical training on psychomotor, affective and cognitive variables during real clinical performance of first year osteopathic medical students.* Unpublished doctoral dissertation, Michigan State University, 1973.

3. Tinning, F. C., Taylor, J. L., & West, D. B. *Analysis of the clinical education program--College of Osteopathic Medicine.* Unpublished manuscript, College of Osteopathic Medicine, Michigan State University, 1975.

4. West, D. B., Markert, R. J., & Bernier, F. A. *Predictors of academic achievement at Michigan State University's College of Osteopathic Medicine.* Occasional Paper No. 11. East Lansing, Mich.: College of Osteopathic Medicine, Michigan State University, 1979.

BIBLIOGRAPHY

BIBLIOGRAPHY

AAMC. *1977-78 AAMC curriculum directory.* Washington, D.C.: Association of American Medical Colleges, 1977.

Bartlett, J. W. Medical school and career performances of medical school students with low Medical College Admission Test scores. *Journal of Medical Education*, 1967, *42*, 231-237.

Best, W. R., Diekema, A. J., Fisher, L. A., & Smith, N. E. Multivariate predictors in selecting medical students. *Journal of Medical Education*, 1971, *46*, 42-50.

Blalock, H. M. *Causal models in the social sciences.* Chicago: Aldine-Atherton, 1971.

Bloom, B. S. (Ed.) *Taxonomy of educational objectives. Handbook 1: Cognitive domain.* New York: Longmans, Green, 1956.

Boldt, R. F. *Factor analysis of law school grades.* (ETS RB-73-42). Princeton, N.J.: Educational Testing Service, 1973.

Buehler, J. A., & Trainer, J. B. Prediction of medical school performance and its relationship to achievement. *Journal of Medical Education*, 1962, *37*, 10-18.

Burt, C. The relations of educational abilities. *British Journal of Educational Psychology*, 1939, *9*, 45-71.

Burt, C. *Three reports on the distribution and relations of educational abilities.* London: King, 1917.

Burt, R. S. Confirmatory factor-analytic structures and the theory construction process. *Sociological Methods and Research*, 1973, *2*, 131-190.

Carroll, J. B. How shall we study individual differences in cognitive abilities?--Methodological and theoretical perspectives. *Intelligence*, 1978, *2*, 87-115.

Cooley, W. W. Who needs general intelligence? In L. B. Resnick (Ed.), *The nature of intelligence.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976, 57-61.

Crowder, D. G. Prediction of first-year grades in a medical college. *Educational and Psychological Measurement*, 1959, *19*, 637-639.

Dawes, R. M., & Corrigan, B. Linear models in decision making. *Psychological Bulletin*, 1974, *81*, 95-106.

Ebel, R. L. Knowledge vs. ability in achievement testing. In *Proceedings of the 1969 invitational conference on testing problems: Toward a theory of achievement measurement*. Princeton, N.J.: Educational Testing Service, 1969, 66-76.

Ebel, R. L. *Measuring educational achievement*. Englewood Cliffs, N.J.: Prentice Hall, 1965.

Elstein, A., Shulman, L. S., Sprafka, S., et al. *Medical problem solving: An analysis of clinical reasoning*. Cambridge, Mass.: Harvard University Press, 1978.

Erdmann, J. B., Mattson, D. E., Hutton, J. G., & Wallace, W. L. The Medical College Admission Test: Past, present, and future. *Journal of Medical Education*, 1971, *46*, 937-946.

Finn, J. D. *A general model for multivariate analysis*. New York: Holt, Rinehart and Winston, 1974.

Flexner, A. *Medical education in the United States and Canada*. A report of the Carnegie Foundation for the Advancement of Teaching, 1910.

Gagné, R. M. Learning and instructional sequence. In F. N. Kerlinger (Ed.), *Review of research in education: I*. Itasca, Ill.: Peacock, 1974, 1-33.

Gaier, E. L. The criterion problem in the prediction of medical school success. *Journal of Applied Psychology*, 1952, *36*, 316-322.

Goldberger, A. S., & Duncan, O. D. (Eds.) *Structural equation models in the social sciences*. New York: Seminar Press, 1973.

Goodman, L. A. A modified multiple regression approach to the analysis of dichotomous variables. *American Sociological Review*, 1972, *37*, 28-46.

Gorsuch, R. L. *Factor analysis*. Philadelphia: W. B. Saunders, 1974.

Gough, H. G. How to select medical students. *Medical Teacher*, 1979, *1*, 17-20.

Gough, H. G. The recruitment and selection of medical students. In R. H. Coombs & C. H. Vincent (Eds.), *Psychological aspects of medical training*. Springfield, Ill.: Charles C. Thomas, 1971.

Gough, H. G. Some predictive implications of premedical scientific competence and preferences. *Journal of Medical Education*, 1978, *53*, 291-300.

Gough, H. G., Hall, W. B., & Harris, R. E. Admissions procedures as forecasters of performance in medical training. *Journal of Medical Education*, 1963, *12*, 983-998.

Gough, H. G., Hall, W. B., & Harris, R. E. Evaluation of performance in medical training. *Journal of Medical Education*, 1964, *39*, 679-692.

Harman, H. *Modern factor analysis* (2nd ed.). Chicago: University of Chicago Press, 1967.

Hauser, R. M., & Goldberger, A. S. The treatment of unobservable variables in path analysis. In H. L. Costner (Ed.), *Sociological methodology 1971*. San Francisco: Jossey-Bass, 1971, 81-117.

Hoffman, E. L., Wing, C. W., & Lief, H. I. Short and long term predictions about medical students. *Journal of Medical Education*, 1963, *38*, 852-857.

Howell, M. A. Personal effectiveness of physicians in a federal health organization. *Journal of Applied Psychology*, 1966, *50*, 451-459.

Howell, M. A., & Vincent, J. W. The Medical College Admission Test as related to achievement tests in medicine and to supervisory evaluations of clinical physicians. *Journal of Medical Education*, 1967, *42*, 1037-1044.

Hyman, H. *Survey design and analysis*. Glencoe, Ill.: Free Press, 1955.

Institute of Medicine. *Cost of education in the health professions* (Interim Report). Washington, D.C.: Institute of Medicine, National Academy of Sciences, 1973.

Jensen, A. R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, *39*, 1-123.

Johnson, D. G. A multi-factor method of evaluating medical school applicants. *Journal of Medical Education*, 1962, *37*, 656-665.

Jöreskog, K. G. Analyzing psychological data by structural analysis of covariance matrices. In D. H. Krantz, R. C. Atkinson, R. D. Luce & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 1). San Francisco: W. H. Freeman & Company, 1974.

Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 1969, *34*, 183-202.

Jöreskog, K. G. Simultaneous factor analysis in several populations. *Psychometrika*, 1971, *36*, 409-426.

Jöreskog, K. G., & Lawley, D. N. New methods in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology*, 1968, *21*, 85-96.

Kegel-Flom, P. Predicting supervisor, peer, and self ratings of intern performance. *Journal of Medical Education*, 1975, *50*, 812-815.

Kerlinger, F. N., & Pedhazur, E. J. *Multiple regression in behavioral research.* New York: Holt, Rinehart & Winston, 1973.

Korman, M., & Stubblefield, R. L. Medical school evaluation and internship performance. *Journal of Medical Education*, 1971, *46*, 670-673.

Kropp, R. P., & Stoker, H. W. *The construction and validation of tests of the cognitive processes described in the "Taxonomy of educational objectives."* (Cooperative Research Project No. 2117). U.S. Office of Education, 1966.

Krupka, J. W., Elstein, A. S., Molidor, J. B., King, L., Parsons, M., & Son, L. *Assessment of empathy skills and problem solving skills as a screen for admission to medical school.* Final report to the National Fund for Medical Education. East Lansing, Mich.: Office of Medical Education Research and Development, Michigan State University, 1977.

Lawley, D. N., & Maxwell, A. E. *Factor analysis as a statistical method* (2nd ed.). London: Butterworths, 1972.

Maatsch, J. L., Downing, S., Sprafka, S., & Holmes, T. Toward a testable theory of physician competence: An experimental analysis of a criterion-referenced specialty certification test library. In *Proceedings of the seventh annual conference on research in medical education*. Washington, D.C.: AAMC, 1978, 399-408.

Maatsch, J. L., & Elstein, A. S., et al. *Model for criterion-referenced medical specialty test.* (Progress Report on Grant No. HS 02038.) East Lansing, Mich.: Office of Medical Education Research and Development, Michigan State University, 1979.

Madaus, G. F., Woods, E. M., & Nuttall, R. L. A casual model of Bloom's taxonomy. *American Educational Research Journal*, 1973, *10*, 253-262.

Markert, R. J. The relationship between grades and clinical competence among first year osteopathic medical students. *Medical Education*, 1978, *12*, 282-286.

McGuire, F. L. Fifteen years of predicting medical student performance. *Journal of Medical Education*, 1977, *52*, 416-417.

Nunnally, J. *Psychometric Theory.* New York: McGraw-Hill, 1967.

Reynolds, H. T. *Analysis of nominal data.* Beverly Hills, Calif.: Sage Publications, 1977.

Rhoads, J. M., Gallemore, J. L., Gianturco, D. T., & Osterhout, S. Motivation, medical school admissions, and student performance. *Journal of Medical Education*, 1974, *49*, 1119-1127.

Richards, J. H., & Taylor, C. W. Predicting academic achievement in a college of medicine from grades, test scores, interviews, and ratings. *Educational and Psychological Measurement*, 1961, *21*, 987-994.

Richards, J. M., Taylor, C. W., & Price, P. B. The prediction of medical intern performance. *Journal of Applied Psychology*, 1962, *46*, 142-146.

Schmidt, F. L. The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement*, 1971, *31*, 699-705.

Schoenfeldt, L. F., & Brush, D. H. Patterns of college grades across curricular areas: Some implications for GPA as a criterion. *American Educational Research Journal*, 1975, *12*, 313-321.

Schumacher, C. F. A factor-analytic study of various criteria of medical student accomplishment. *Journal of Medical Education*, 1964, *39*, 192-195.

Simon, H. A. *Models of man.* New York: Wiley, 1957.

Sirotkin, R. A., & Whitten, C. F. Relationships between the preclinical and clinical years of medical school: A study of the interrelatedness of several performance measures. In *Proceedings of the seventeenth annual conference on research in medical education.* Washington, D.C.: AAMC, 1978, 197-201.

Sörbom, D., & Jöreskog, K. G. *COFAMM: Confirmatory factor analysis with model modification.* User's Guide. Chicago: National Educational Resources, 1976.

Sörbom, D., & Jöreskog, K. G. *LISREL: Analysis of linear structural relationships by the method of maximum likelihood.* User's Guide. Chicago: National Educational Resources, 1978.

Spearman, C. *The abilities of man.* London: Macmillan, 1927.

Spearman, C. General intelligence objectively determined and measured. *American Journal of Psychology*, 1904, *15*, 201-293.

Stewart, D., & Love, W. A general canonical correlation index. *Psychological Bulletin*, 1968, *70*, 160-163.

Thurstone, L. L. *Multiple-factor analysis.* Chicago: University of Chicago Press, 1947.

Thurstone, L. L. Primary mental abilities. *Psychometric Monographs*, 1938, *1*.

Tucker, L. R., & Lewis, C. A reliability coefficient for maximum-likelihood factor analysis. *Psychometrika*, 1973, *38*, 1-10.

Vernon, P. E. *The structure of human abilities* (2nd ed.). New York: Wiley, 1961.

Wang, M. C., & Stanley, J. C. Differential weighting: A review of methods and empirical studies. *Review of Educational Research*, 1970, *40*, 663-705.

Wiley, D. E., Schmidt, W. H., & Bramble, W. J. Studies of a class of covariance structure models. *Journal of the American Statistical Association*, 1973, *68*, 317-323.

Wingard, J. R., & Williamson, J. W. Grades as predictors of physicians career performance: An evaluative literature review. *Journal of Medical Education*, 1973, *48*, 311-322.

Wolfle, D. Factor analysis to 1940. *Psychometric Monographs*, 1940, *3*.