PREDICTIVE VALIDITY, DIFFERENTIAL VALIDITY, AND DIFFERENTIAL PREDICTION OF THE SUBTESTS OF THE MEDICAL COLLEGE ADMISSION TEST

A dissertation presented to

the faculty of the College of Education of

Ohio University

In partial fulfillment

of the requirements for the degree

Doctor of Philosophy

Lydia S. Kyei-Blankson

June 2005

This dissertation entitled

PREDICTIVE VALIDITY, DIFFERENTIAL VALIDITY, AND DIFFERENTIAL PREDICTION OF THE SUBTESTS OF THE MEDICAL COLLEGE ADMISSION TEST

BY

LYDIA S. KYEI-BLANKSON

has been approved for

the Department of Educational Studies

and the College of Education by

George Johanson

Professor of Educational Research and Evaluation

James Heap

Dean, the College of Education

KYEI-BLANKSON, LYDIA Ph.D. June 2005. Educational Research and Evaluation

Predictive Validity, Differential Validity, and Differential Prediction of the Subtests of the Medical College Admission Test (298pp.)

Director of Dissertation: George Johanson

MCAT scores and undergraduate GPA are two screening devices that are widely used by medical schools in the U.S. to select aspiring medical students. Given their widespread use, it is important that the validities of MCAT scores and undergraduate GPA be investigated to ensure the accuracy of medical school admission decisions.

This study was therefore designed to address three important aspects of validation, namely, predictive validity of MCAT subtest scores and undergraduate GPA, differential validity, and differential prediction of the MCAT subtest scores. First, predictive validity was evaluated as an index of the relationship between the predictors, MCAT subtest scores and undergraduate GPA, and the criterion, first-year medical school GPA. Second, differential validity was assessed by comparing the magnitude of validity coefficients obtained as the correlation between first-year medical school GPA and MCAT subtest scores for men and women and for White, Black, Asian, and Hispanic medical students. Third, differential prediction was examined by testing for differences in the regression systems obtained for the different subgroups of examinees.

The sample used in the study consisted of 3,187 students drawn from 1992 and 1993 cohorts of 14 medical schools. Statistical procedures utilized in the research included regression analysis, Fisher's z transformations, F-ratio test of equality of standard errors of estimate, and ANCOVA tests of equality of regression slopes and intercepts.

Results obtained showed moderately high correlations between the predictors and the criterion. MCAT subtest scores and previous grades were individually good predictors of medical school freshman grades. The combination of MCAT subtest scores and undergraduate GPA was, however, a more powerful indicator of performance in the first-year of medical school.

Differential validity results showed that in most cases women had higher validity coefficients compared to men. With regards to differential prediction, the results implied that using common regression equations derived from a pooled sample of examinees to predict performance in the first year of medical school may result in underprediction for Whites and overprediction for Blacks and Hispanics.

Approved:

George Johanson

Professor of Educational Research and Evaluation

Acknowledgements

This study would not have been possible without the help of many people. I gratefully acknowledge all the people who have given me support through out my study and through out my life.

First and foremost, l am deeply indebted to Dr. Patricia Etienne, Director of MCAT Research for suggesting this topic. Also, many thanks go to Su Kotehne of the MCAT Research Division for providing me with the necessary data to make this study possible.

I acknowledge my advisor, Dr. George Johanson, for his guidance through out this research. I am eternally grateful for his generous sharing of his invaluable knowledge and time to help me with this dissertation.

I am also appreciative of my dissertation committee members, Dr. Teresa Franklin, Dr. Gordon Brooks, and Dr. Gary Moden, who provided intellectual support, constructive criticisms, and invaluable suggestions that significantly improved the content of this dissertation.

I extend my sincere gratitude to my extended and immediate family, particularly my husband, Dr. Joseph Blankson, for his unfailing support; to my children, Kweku, Kwame, and Ama, who serve as my source of inspiration in life; and to my parents, Mr. and Mrs. James Kyei, and my brothers and sisters for their constant encouragement.

Last but certainly not least, my thanks go to my friends, Mr. and Mrs. Kwesi Boateng, Dr. and Mrs. Kwesi Tandoh, Dr. and Mrs. Anthony Owusu-Ansah, and Afua Nyarko, who have unselfishly supported my family in diverse ways, especially caring for my children during the times that l had to study. I thank God for all of you.

Table of Contents

List of Tables

Chapter One: Introduction

Admission to medical school is typically extremely competitive. Each year the 125 medical schools in the United States receive thousands of applications from a large number of aspiring candidates. In the 2002-2003 academic year, 33,625 applicants submitted 373,686 applications for the 17,592 available slots (Brazansky & Etzel, 2003). The average applicant files approximately eleven applications (Mitchell, Haynes, & Koenig, 1994).

Medical school admissions committees use a set of criteria, which generally include students' undergraduate grade point average (GPA), medical-related work experience, extra-curricular activities and interests, letters of evaluation from faculty members, and interview ratings (Elam, Seaver, Berres, & Brandt, 2000; Lakhan, 2003), to identify the most competitive students out of their applicant pool for admission. Another important criteria that is required by almost all medical schools as a prerequisite to admission is the Medical College Admission Test (MCAT).

The MCAT is a standardized examination that consists of both multiple-choice and essay items. The Association of American Medical Colleges (AAMC) administers this examination twice annually. AAMC developed the MCAT with the intent to assess candidates' understanding of science prerequisites necessary for the study of medicine; provide a mechanism for the evaluation of analytical skills; and provide a basis by which medical admissions committees can attempt to predict performance in a medical curriculum (Mitchell, 1987).

Historical Background of the MCAT

The MCAT, originally known as the Scholastic Aptitude Tests for Medical Schools, was created in 1928. Before then, the medical school screening process was uneven. Evaluation of medical school applicants was based on college grades from two to four years of undergraduate education or sometimes merely on a high school diploma, supplemented with biographical information and letters of endorsement (McGaghie, 2002a). Inconsistencies in admissions standards led to the acceptance of students who were intellectually unprepared to tackle the medical curriculum. Such students were later on faced with academic difficulty, resulting in their attrition from the medical program. By the early twentieth century, medical colleges were battling attrition rates as high as fifty percent (Moss, 1930).

High attrition rates represented a huge waste of human capital, individual and family aspirations, faculty time and energy, and misspent tuition money (McGaghie, 2002a). It became evident therefore that admissions committees needed to come up with a standardized measure to better evaluate applicants' readiness for medical education (McGaghie, 2002a; Sedlacek, 1967). This led to the introduction of the MCAT.

The MCAT has undergone five revisions since its introduction in 1928. The 1928 Scholastic Aptitude Tests for Medical Schools consisted of multiple-choice and true-false questions organized into six to eight subtests that focused on memory, knowledge of scientific terminology, reading and comprehension, and logic (McGaghie, 2002a). The Scholastic Aptitude Tests for Medical Schools was renamed the Professional Aptitude Test in 1946, and again the Medical College Admission Test in 1948. The 1946 Professional Aptitude Test contained only four subtests: Verbal Ability, Quantitative

Ability, Science, and Understanding Modern Society; the test remained identical in content and structure in 1948 (Sedlacek, 1967). In 1962, the Understanding Modern Society section was reconceived and expanded to a broader test known as General Information. However, in 1977, the General Information subtest was eliminated, and the Verbal, Quantitative, and Science sections of the test were modified (Littlemeyer & Mauney, 1977).

The current version of the test, the 1991 MCAT, places greater weight on breadth of academic background, reasoning skills, and writing ability of future physicians (Holden, 1989) and was designed to meet rapid changes and expansions in the knowledge base and technologies of medicine (Anderson & Swanson, 1993; Swanson & Mitchell, 1989). The new changes required that, generally, physicians must be able to gather and assess data, apply the basic concepts and principles of medicine to solve scientific and clinical problems, continually update their knowledge and skills, and communicate effectively with patients, colleagues, and the public (Mitchell, Haynes, & Koenig, 1994).

The 1991 MCAT is made up of four sections, namely, Biological Sciences, Physical Sciences, Verbal Reasoning, and Writing Sample. These four areas assess mastery of basic concepts in biology, chemistry, and physics; facility with scientific problem solving and critical thinking; and communication and writing skills (Mitchell, 1991).

The Biological Sciences, Physical Sciences, and Verbal Reasoning components of the MCAT are comprised of multiple-choice questions. The scores on these sections currently range from 1 (*lowest*) to 15 (*highest*) (Medical College Admission Test, 2003).

Until 2002, however, the scores on the verbal test ranged from 1 (*lowest*) to 13-15 (*highest*) (Association of American Medical Colleges, 2003).

The Biological Sciences section tests examinees' reasoning in biology and organic chemistry (Medical College Admission Test, 2003). It consists of a number of passages followed by approximately eight questions per passage. It also includes questions that are independent of the passages and of each other.

The format of the Physical Sciences section is similar to that of the Biological Sciences subtest in that it also includes passages with approximately eight questions per passage. The Physical Sciences section assesses candidates' reasoning skills in introductory general chemistry and physics (Medical College Admission Test, 2003).

The Verbal Reasoning section, on the other hand, contains questions from the areas of humanities, social sciences, and natural sciences. The section includes several lengthy passages, which are accompanied by about ten questions per passage. The intent of this subtest is to assess examinees' ability to understand, evaluate, and apply information and arguments presented in prose texts (Mitchell, 1991).

The Writing Sample is the newly added section of the MCAT that requires candidates to develop and present ideas coherently (Gilbert, Basco, Blue, & O'Sullivan, 2002) and provides evidence of examinees' analytic thinking and writing skills, which are considered critical to the preparation of useful medical records and to effective communication with patients and other health professionals (Mitchell, 1991). This subtest consists of two 30-minute essays. Two separate readers rate the essays; essays receiving scores that differ by more than one point are then re-evaluated by a third reader who determines the total score. The total score for an examinee's essays are then converted to an alphabetical scale for reporting (Hojat, Erdmann, Veloski, Nasca, Callahan, Julian, et al. 2000). Scores on the Writing Sample range from 1 (*lowest*) to 11 (*highest*) and are reported alphabetically from J to T, respectively (See rubrics in Appendix H).

Since the introduction of the MCAT, the AAMC has expended considerable effort in assessing its validity, especially its ability to predict success in medical school (Gough, Hall, & Harris, 1963; Julian & Lockwood, 2000; Koenig, Huff, & Julian, 2002). In the *Standards for Educational and Psychological Testing,* validity is referred to as " the most important consideration in test evaluation" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1985, p. 9); and validation is an important continuous process that is crucial for the accurate use of tests.

Test Validity

The degree of validity is the single most important aspect of any test (Mehrens & Lehmann, 1987). Validity refers to "the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores" (American Educational Research Association, American Psychological Association, & National Council of Measurement in Education, 1985, p. 9).

One form of validity that is viewed as essential for admission testing is predictive validity. Establishing the predictive validity of scores on an instrument is at issue when the purpose of the instrument is to estimate performance on a criterion measure (Anastasi, 1988). Rudner (1994) explained that predictive validity is demonstrated when inferences regarding achievement are established via a statistical relationship between test scores and subsequent academic performance. In other words, empirical evidence in support of predictive validity includes a comparison of performance on a test against performance on an external measure.

Although the detection of other validity components is important for any given test, this study focuses on only predictive validity because according to Linn (1990), predictive validity has long been a very important part of the evidence provided to support the use of tests in admissions.

Predictive Validity of the MCAT

The predictive validity of each version of the MCAT is assessed to provide admissions officials with information on the usefulness of the MCAT scores for predicting performance in medical school (Mitchell, 1991). It has been asserted that the old versions of the MCAT adequately fulfilled their intended function of providing highly dependable measures of students' ability and achievement in the medical curriculum (Sanazaro & Hutchins, 1963).

The predictive validity of the current MCAT was estimated prior to its introduction in 1991 using preadmission information such as undergraduate GPA, undergraduate institutional selectivity, and pre-1991 MCAT scores of medical school entrants to sixteen institutions. Based on the validity coefficients reported with respect to medical school grades at the sixteen medical institutions, it was confirmed that the MCAT was useful for predicting students' performance in medical school (Mitchell, 1991).

Since then, a number of studies have investigated the predictive validity of the 1991 MCAT. In almost all the studies, researchers employ students' undergraduate GPA as an additional predictor and the results of the studies consistently suggest that MCAT scores and undergraduate GPA, individually and in combination, contribute significantly to the prediction of medical school grades.

Problem Statement

Medical school admissions committees are charged to ensure that candidates for admission are academically fit for medical school (Shepard, 1987). Generally, acceptance into medical school is based on factors such as letters of recommendation, performance on a personal interview, undergraduate GPA, and MCAT scores.

MCAT scores and students' undergraduate GPA are considered to be the most important criteria for the selection of aspiring medical students (Mitchell, Haynes, and Koenig, 1994). Medical school admissions committees use MCAT scores to corroborate their judgment of applicants' intellectual ability, which is determined from candidates' undergraduate GPA. The intelligent use of these two sources of information, therefore, can greatly strengthen the medical school admissions process. For this reason, empirical studies that examine their predictive validity are necessary to help admissions committees make sound evidence-based decisions (Sedlacek, 1967).

At this point, it is important to draw attention to the fact that the use of MCAT scores for making admissions decisions varies from committee to committee. For instance, some admissions committees consider applicants' scores on the individual subtests of the MCAT while others are only interested in their applicants' total score on the MCAT. However, after using keywords such as "medical college admission test", "predictive validity", "college entrance examination", "medical education", and "admission criteria" to locate past MCAT predictive validity studies via article databases like Education Abstracts, Educational Resources Information Center (ERIC), Dissertation

Abstracts, Medline, and PsycInfo, it was observed that the majority of the studies used applicants' total score on the MCAT as the predictor in their analyses. Erdmann (1977) questioned the use of such methodology by stating that simply adding the subtest scores to obtain a total MCAT score suggests that the individual subtests of the MCAT equally predict students' performance when in fact they may not.

Another observation was that validity evidence gathered in most of the previous studies was based on analysis of aggregate student data. Considerably fewer studies have investigated whether MCAT scores predict performance equally across examinees' subgroups such as gender and ethnicity (Koenig, Sireci, & Wiley, 1998). Thus, the predictive strength of the MCAT is less clear when students' gender and ethnicity are considered (Veloski, Callahan, Xu, Hojat, & Nash, 2000).

The objective of this study, therefore, is to examine the predictive validity of the individual subtests of the MCAT. The study also aims at determining whether the MCAT subtests have equal predictive strength across applicants' gender and ethnicity. The individual MCAT subtests scores will be tested for predictive validity and then tested for differential validity and differential prediction across gender and ethnicity.

In this study, undergraduate GPA is included as an additional predictor. The advice to use information other than test scores is based, in part, on the recognition that undergraduate GPA plays a major role in the admissions process; previous academic records not only improve the prediction of future academic performance that can be achieved with test scores alone, but often the two predictors in combination yield better prediction than either one alone (Linn, 1990).

Since the MCAT is designed to help identify candidates who have the prerequisite science, verbal, and writing skills necessary for success in medical school, the criterion measure employed in this study is medical school GPA rather than future physician skills; first-year medical school GPA is the specific criterion used in the study. This criterion is considered because studies at the graduate school level follow a parallel pattern with first-year grades as the criterion (Linn, 1990). Also, according to Wightman (2003), first-year GPA is most often used as the criterion measure in predictive validity studies because they become available within a year of the start of school and are based on a composite of academic performance accumulated over a year of school, thus allowing for differences in course difficulty to average out; most of the core courses for a program are taken during the first year of school, and therefore, the content on which GPA is based tends to be more consistent across students than it is at any later time.

Research Questions

The study is focused on the following research questions:

1. How well do undergraduate GPA and MCAT subtest scores, both individually and collectively, predict first-year medical school GPA?
2. How well do MCAT total/composite scores predict first-year medical school GPA?
3. How well do subsets of MCAT subtest scores predict first-year medical school GPA?
4. Are the predictive strengths of MCAT subtest scores, both individually and collectively, consistent across gender?

5. Are the predictive strengths of MCAT subtest scores, both individually and collectively, consistent across ethnicity?

Significance of the Study

A central issue in every admissions process is the selection of candidates who are most likely to succeed in a curriculum. The information on which such selections are based should therefore be accurate and reliable in order to ensure the integrity of such decisions.

Most colleges use standardized test scores as a criterion for admissions. The scores on such tests demonstrate applicants' intellectual ability and knowledge in their desired fields of study. They also aid in the prediction of applicants' success or failure in a program of study.

Assessing test validity is one way to ensure that the information gathered from such test scores is accurate. Test validity is therefore considered to be the most fundamental and important in psychometrics (Angoff, 1988) and an understanding of this concept is in itself the foundation for fair and proper use of tests and measurements of all kinds (Ebel & Frisbie, 1991). One type of validity, predictive validity, is viewed as very important in situations where tests are used in making admissions decisions (Nunnally, 1978).

The aim of the medical school admissions committees is to select students who are most likely to perform well in the medical curriculum. Success in the medical curriculum requires a certain level of intellectual ability and knowledge in various aspects relevant to medical education. The MCAT provides such information on applicants (Sedlacek, 1967) and ensures the selection and retention of students who are "ready" for medical school. Thus, studies focusing on the validity of the MCAT are of great importance to medical schools.

Over the years a large number of studies have evaluated the content, criterion, and construct validity of the MCAT. So far, the majority of studies have evaluated the predictive validity of the MCAT, the ability of the MCAT to predict success in medical school. Most often, undergraduate GPA is included as an additional predictor in these studies. The results have shown that the MCAT provides indication of students' performance in medical school, and support the use of MCAT scores for making medical school selection decisions.

Although the information provided by such studies is very useful for admissions committees, it should be noted that the results from these studies were mostly obtained from analyses of aggregate MCAT scores and aggregate students' data. Studies dealing with the predictive validity of the scores on the individual MCAT subtests, as well as the predictive strength of the subtest scores across the various groups of examinees are very few (Koenig, Sireci, & Wiley, 1998; Jones & Vanyur, 1985).

Cullen, Peckham, and Schwarz (1980) mentioned that studies to estimate the predictive validity of the subtests of the MCAT are practical for admissions committees, especially those who consider scores on the MCAT subtests as indicators of preparation for and success in medical school (Stefanu & Farmer, 1971; Mitchell, Haynes, & Koenig, 1994). Also, such studies are significant because each subtest provides separate information on the achievement levels and the strengths and weaknesses of applicants (Mitchell, 1991).

The MCAT is administered to examinees from diverse backgrounds and therefore, the validity of score interpretations for one group of individuals may not be the same as for another group (Linn & Hastings, 1984). Since test scores and previous grades are used to predict students' future academic achievement, it is important to know whether the validities differ for the different identifiable subgroups of applicants (Linn, 1990); this concept has relevance for issues of test bias and fair test use and is important in test validation (Young, 2001). Such information will help admissions committees to gain a clearer understanding of potential differences in MCAT subtest scores across examinees' subgroups and have implications for the estimation of separate prediction equation models for the different subgroups to ensure equitable selection in medical school (Crocker & Algina, 1986).

Delimitations and Limitations of the Study

The delimitation of the study is that it examines the predictive validity of the MCAT subtest scores. The study also investigates whether the predictive power of each MCAT subtest differs across examinees' gender and ethnicity. The data used in the study includes information on MCAT subtest scores, undergraduate GPA, and first-year medical school GPA for 1992 and 1993 matriculants to 14 medical schools. The schools were selected in order to obtain a representative sample of the 125 medical schools in the U.S.

The study is limited in that it focuses on predictive validity, which is only one aspect of test validity. Another limitation is that only one other admission criteria, undergraduate GPA, is considered in the study. Other admission criteria, such as interview ratings, letters of recommendation, and medical-related work experience, which are typically employed in the medical school admissions process, are not included in the study. Also, with regards to the data used in the study, undergraduate GPA, which consists of both science and non-science GPA, is reported on a 4-point scale for all of the participants of the study. There is no indication as to whether the 4-point scale actually applies to all the undergraduate institutions from which the GPAs were obtained. The scores on the subtests of the MCAT are reported as scaled scores as well with very limited information on how the scaled scores were obtained.

Definition of Terms

Differential Prediction

A finding where the prediction equations and/or standard errors of estimate obtained from analyses are significantly different for different groups of examinees (Young, 2001). Such a finding indicates that the future performance of different subgroups, for example men and women, cannot be predicted using a single prediction equation. Rather, there is a need to derive different prediction equations for the different subgroups of applicants.

Differential Validity

A finding where the computed validity coefficients obtained for the subgroups of examinees are significantly different (Young, 2001). In other words, the predictive validity of the test scores, which is measured as the relationship between predictors such as preadmission variables and the criterion measure or future performance, is significantly different for the different subgroups of examinees.

First-year Medical School GPA

In this study, first-year medical school GPA refers to the average for all grades obtained from all courses taken by the end of the first year of medical school. The resulting averages are reported on a scale of 0 to 4 or 0 to 100 for the cohorts from the 14 medical schools included in the study. Generally, courses taken by the end of the first year include basic science courses such as Gross Anatomy, Biochemistry, Physiology, Medical Genetics, Behavioral Sciences, and Physical Diagnosis. Classes are typically conducted in modules using full lectures, small-group discussions, seminars, workshops, and laboratory teachings; grades obtained from these classes are based almost entirely on written, oral, and lab exams (Wilkinson, 2004). Although the grading system varies from school to school, the majority of medical schools use the standard A/B/C/D/F scale. Some medical schools, however, employ an ABCF scale, a Pass/Pass*/Fail scale, or an Honors/High Pass/Pass*/Fail scale (Association of American Medical Colleges, 2003).

MCAT Subtest Scores

This refers to the individual scores on the Biological Sciences, Physical Sciences, Verbal Reasoning, and Writing Sample subtests of the MCAT. The Biological Sciences and Physical Sciences subtests focus on basic biology, chemistry, and physics concepts and scaled scores on these subtests are reported on a 15-point scale. The Verbal Reasoning subtest focuses on verbal skills with reported scales ranging from 1 (*lowest*) to 13-15 (*highest*). Lastly, the Writing Sample subtest focuses on written skills and scores on this test are reported on a scale of 1-11.

Predictive Validity

Predictive validity indicates how accurately test data can predict criterion scores obtained at a later time (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999). Predictive validity provides an indication of the ability of a test to predict future performance or behavior.

Undergraduate GPA

Undergraduate GPA is the weighted average of all grades obtained from courses taken at the undergraduate level. In this study, undergraduate GPA consists of undergraduate science GPA and non-science GPA and both are reported on a 4-point scale.

Organization of the Study

The material in this dissertation is organized into five chapters: Introduction, Literature Review, Methodology, Results, and Summary, Interpretation, Conclusions and Recommendations. A list of the references cited in the dissertation is subsequently provided.

The introduction provides a background to the study. This chapter also focuses on the problem statement, research questions, significance of the study, and delimitations and limitations of the study.

In the chapter two, the literature review, an overview of previous research relating to predictive validity of the MCAT is presented. The chapter also focuses on other topics relevant to the study. Such topics include a historical overview of higher education admissions criteria in the U.S. and the effectiveness of GPA and standardized test scores as selection criteria in the admissions process. The chapter also discusses medical school admissions and the effectiveness of MCAT and undergraduate GPA as medical school admissions criteria.

Chapter three provides a detailed description of the instrumentation, research design, data source, sample, data collection procedures, and data analysis procedures employed in the study.

Finally, results of the analysis are reported in chapter four and the last chapter, chapter five, provides a summary of the dissertation, interpretation of the research findings, conclusions, and recommendations for admission committees and for future research.

Chapter Two: Literature Review

This chapter provides a review of existing literature related to the domain of this study. First, the chapter presents a historical overview of college admission requirements in the U.S. Then, two of the most widely used criteria for admission, GPA and standardized tests, are discussed. The chapter also provides an overview of the medical school admissions criteria and later addresses the effectiveness of GPA and MCAT scores as predictors of medical school performance.

Historical Overview of College Admission Requirements in the U.S.

Each year millions of students apply for admission to various programs in institutions of higher education. In most cases, the number of applicants exceeds the number of available positions. As a result, admissions committees use a set of requirements to identify the most competitive applicants for admission. Admission requirements or criteria are indices assumed to be a measure of future academic performance or success and are used either implicitly or explicitly to select applicants for admission to undergraduate, graduate, or professional institutions (Hirschberg, 1977).

The history of college admission requirements in the U.S. can be traced to 1642 when decisions to admit students to Harvard College were based on their ability to construe and grammatically resolve Greek and Latin. Other requirements for admission included academic records, students' moral conduct, and performance on an interview with the president or senior tutor of the institution (Broome, 1903). By the middle of the eighteenth century, college admission requirements expanded to include knowledge of elementary arithmetic and the ability to transcribe English into Latin.

A demand for less classical and more practical entry requirements, in the nineteenth century, led to the rise of academies that emphasized a wider range of subjects such as English, science, geography, mathematics, and history, and to some extent music, bookkeeping, and other vocational subjects, for admission (Fine, 1946). Gradually, college admission requirements became more flexible and less uniform, varying from school to school. Concern for the lack of uniformity led to discussions on standardizing admission requirements for all colleges (Westoff, 1980).

By the beginning of the twentieth century, some form of uniformity was attained. The majority of the colleges and universities required completion of a four-year high school for undergraduates or completion of a four-year baccalaureate for graduate and professional students, letters of recommendation, personal interviews, GPA, and scores on entrance tests, which were developed and controlled by the individual colleges (Kingsley, 1913; Fine, 1946; Traxler & Townsend, 1953).

The individual college entrance tests were later replaced with standardized tests, which arguably did a better job at demonstrating students' readiness for enrollment (Bowles, 1956). The first testing program to be administered on a large scale for the purpose of college admission was the Scholastic Aptitude Test (Angoff, 1971), which is now popularly known as SAT. A follow-up study by Fishman and Pasanella (1960) and results from a survey of 250 colleges conducted by Berger (1961) revealed that this standardized test had become an acceptable and absolutely essential requirement in the college admissions process.

In the latter part of the twentieth century, the selection of students based on predicted performance using mostly mathematical formulas and models based on test scores became quite evident (Fincher, 1992). In the same period, colleges and universities began emphasizing the importance of non-cognitive abilities to the admissions process (Beale, 1970; Mulvenon, Stegman, Thorn, & Thomas, 1999).

To date, two selection tools, GPA and test scores, remain prime factors in the admissions process. The existing literature consistently identifies GPA and standardized test scores, individually and in combination, as good predictors of students' subsequent performance (Wightman, 2003).

GPA and Standardized Test Scores as Criteria for Admissions

GPA and Admissions

Willingham (1974) indicated that GPA is a principal criterion that is used in the admissions process because it is readily available, quantifiable, equitable, and assumed to be fair. Also GPA presumably measures desired behaviors like intelligence, aptitude, and achievement that are required for students' subsequent studies (Gottheil & Michael, 1957; Hirschberg, 1977; Humphreys, 1962).

The use of GPA as a requirement for admissions has, however, been subjected to a number of criticisms. Anderhalter (1962) and Juola (1968) were primarily concerned with the fact that GPA originated from different colleges and universities that have different grading standards, thus lacks comparability from student to student; similar grades may mean different things at different schools. Linn (1982) later pointed out that the lack of comparability of grades from one college to another and from one curriculum to another is a potentially important source of unfairness. The use of GPA in admissions thus results in systematic bias against certain students, especially those enrolled in the more difficult curricula. Other arguments suggest that GPA is an inflated subjective measure of students' performance. Willingham (1974) stated that GPA is an inflated and less trustworthy measure of students' abilities because it is based on the subjective impressions of faculty.

Standardized Test Scores and Admissions

Traditionally, the purpose of tests has been to measure differences between individuals or between different reactions of the same individual (Aiken, 1994). In the field of education, standardized tests were first primarily used to classify students into different educational tracks with reference to their ability, identify intellectually retarded students on one hand and gifted students on the other, determine promotion from one grade level to the next, diagnose academic failures, and award high school diplomas (Anastasi, 1988; Linn, 2001). According to Linn, the use of standardized tests later expanded to include selection of applicants for college admissions.

Standardized tests are used to make informed decisions about students' admissions to undergraduate, graduate, and professional education (Ferguson, 1975). The basic rationale for using standardized tests in the admissions process is that they are relevant and reliable (Ebel, 1978). They provide admissions committees with a standardized measure of academic achievement for all examinees (Ebel, & Frisbie, 1991).

Nearly all colleges and universities in the U.S. require that applicants take at least one major standardized test prior to admission. The SAT for undergraduates and the Graduate Record Examination (GRE) for graduate students are among the most widely used standardized tests by a large number of colleges. Scores on these tests are most often used to predict students' future achievement in a department's curriculum.

There are, however, a number of criticisms against tests and their use in academic admissions. Some opponents of tests and admissions testing argue that tests measure only examinees' cognitive skills and also that the reliability and validity of tests are not sufficient to justify their use in selection (Wigdor & Garner, 1982).

GPA and Standardized Test Scores as Predictors of Students' Subsequent Performance

Over the years, a sizable amount of research has been conducted on GPAs and standardized test scores and their ability to predict students' subsequent performance. For example, Wilson (1983) reviewed a number of SAT studies in which the test scores and GPA for about 12,000 college students who graduated from 40 institutions between 1930 and 1980 were analyzed. His findings supported the use of GPA and SAT scores in admission.

Burton and Ramist (2001) later reviewed SAT predictive validity studies from 1980 to 2001. The studies covered 80,000 students from 80 institutions. Predictors in the studies included overall SAT scores, SAT Verbal, Mathematics, and Subject Test scores, and high school GPA. Measures of success included cumulative college GPA and first-year college GPA. The results of the study indicated that SAT scores and high school GPA made substantial contributions to the prediction of overall and first-year college GPA, and the combination of SAT test scores and high school GPA provided better validity estimates than either predictor alone.

In 1974, Willingham reviewed 43 GRE validity studies published between 1952 and 1972. Summarizing the results from these studies, he concluded that GRE and undergraduate GPA were valid predictors of graduate GPA. The study further showed that GRE is usually a better predictor than undergraduate GPA and for various fields of study, the GRE Advanced test was the strongest predictor of success in graduate school. Since then, various researchers have indicated that GRE scores are reasonable predictors of graduate school cumulative grade performance (Dollinger, 1989; House, Gupta, and Xiao, 1997).

Generalizations made from these and other validity studies imply that GPA and standardized test scores, across the various testing programs, are effective predictors of future success in higher education (Grandy, 1994; Linn, 1990; Mulvenon, Stegman, Thorn, & Thomas, 1999). However, there are still ongoing discussions regarding the effectiveness of tests, especially the reliability and validity of test scores.

Test Reliability

Reliability is defined as "the consistency of measurements when the testing procedure is repeated on a population of individuals or groups" (American Educational Research Association, American Psychological Association, & National Council of Measurement in Education, 1999, p. 25). Crocker and Algina (1986) also described reliability as the degree to which individuals' deviation scores remain relatively consistent over repeated administration of the same test or alternate test forms. In summary, reliability concerns consistency or stability (Messick, 2000).

Reliability is a property of a set of test scores and not a property of the test itself (Ebel & Frisbie, 1991). There are, traditionally, two types of reliability, namely, reproducibility and internal consistency (Shea & Fortna, 2002). Reproducibility addresses the extent to which examinees' scores on two alternate forms of a test (alternate form reliability) and examinees' scores on the same test administered on different occasions (test-retest reliability) is similar. Internal consistency, on the other hand, is the extent to which a test assesses similar characteristics across examinees.

Reliability is quantified via two indices: standard error of measurement and reliability coefficient. The standard error summarizes potential within-person inconsistency in score-scale units (Feldt & Brennan, 1989) and reliability coefficient measures the amount of error associated with a set of test scores (Ebel & Frisbie, 1991).

Test Validity

Validity refers to the accuracy with which a set of test scores measure a particular cognitive ability of interest (Ebel & Frisbie, 1991). A measurement is valid when it measures what it is supposed to measure and performs the functions that it purports to perform. It is important to note that validity does not refer to the characteristic of a test but rather to the evidence gathered from the test score. Cronbach (1971) argued that an instrument cannot be validated but rather what needs to be valid is the meaning or interpretation of the scores on the instrument, as well as any implications for action that this meaning entails. It is also important to note that validity is a matter of degree, not all or none (Messick, 1989).

In the *Standards for Educational and Psychological Testing*, validity is defined as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of the test" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 9). According to Messick (1989), assessing a test's validity involves an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on the test scores. Simply speaking, test validity is associated with the process of accumulating evidence to support the inferences made from test scores (Wainer & Braun, 1988).

There are three traditional approaches to validity assessment; namely, content, criterion, and construct validity.

Content Validity

Content validity refers to the degree to which samples of items, tasks, or questions on a test represent some defined universe or domain of content (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). It is used in situations where the test user desires to draw an inference from the examinee's test score to a larger domain of items similar to those on the test itself (Crocker & Algina, 1986).

According to Messick (1989), content validity is evaluated by showing how well the content of the test samples the class of situations or subject matter about which conclusions are drawn by obtaining professional judgments about the relevance of the test content of a particular behavioral domain of interest and about the representativeness with which item or task content covers that domain.

Criterion Validity

Criterion validity is at times referred to as criterion-related validity. A criterion measure is an accepted standard against which some test is compared to validate the use of the test's score as a predictor (Ebel & Frisbie, 1991).

Criterion validity is used for situations where the test user desires to draw an inference from the examinee's test score to their score on some behavioral variable of practical importance (Crocker & Algina, 1986). Test scores are, thus, systematically related to one or more criteria that are external to the test.

Criterion-related validity is evaluated by comparing the test scores with one or more external variables considered to provide a direct measure of the characteristic or behavior in question and is based on the degree of empirical relationships, usually in terms of correlations or regressions, between the test scores and criterion scores (Messick, 1989). There are two major types of criterion validity; namely, concurrent and predictive validity.

Concurrent validity indicates the relationship between test scores and outcome criteria that are analyzed at the time that the test was given (Crocker & Algina, 1986). In other words, concurrent validity is a measure of the extent to which test scores estimate an individual's present standing on the criterion (Messick, 1989). This form of validity is evaluated by comparing the test's score to scores on another test taken at the same time. The aim of concurrent validity is to evaluate the effectiveness of the test scores (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

Predictive validity evidence indicates how well an assessment can predict scores obtained at a later time through the use of either the same measure or a different measure. In the *Standards for Educational and Psychological Testing*, predictive validity is defined as "how accurately test data can predict criterion scores that are obtained at a later time" (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999, p. 180). Predictive validity comes into play when a test is used to predict the likelihood of some future performance. It indicates the extent to which an individual's future level on the criterion is predicted from prior test performance (Messick, 1989; Crocker & Algina, 1986).

Construct Validity

The term construct refers to a psychological construct; a theoretical conceptualization about an aspect of human behavior that cannot be measured or observed directly (Ebel & Frisbie, 1991). Construct validity is concerned with the interpretation of a test's score as a measure of some attribute and it involves the collection of empirical evidence to support the theoretical construct that underlies measurement and the resulting inferences (Cronbach & Meehl, 1955). Wainer and Braun (1988) also described construct validity as a process, not a procedure; and it requires many lines of evidence, not all of them quantitative.

Construct validation is used for situations where no criterion or universe of content is accepted as entirely adequate to define the quality to be measured (Messick, 1995). It is evaluated by investigating what qualities a test measures, that is, by determining the degree to which certain concepts or constructs account for performance on the test and is based on an integration of evidence that bears on the interpretation or meaning of the test score (Messick, 1989). Messick further on explained that any kind of information about the test can contribute to the understanding of its construct validity.

Researchers have identified some factors that may affect the results obtained in studies on the validity of standardized test scores. These factors include restriction of range and criterion unreliability and will be discussed in t he next section.

Factors Influencing the Validity of Standardized Tests

Restriction of Range

Givner and Hynes (1979) stated that the range of test scores and grades used in predictive validity studies are typically restricted because these scores are taken from students selected for admission and thus do not reflect the scores of the total population or the total number of candidates for admission. According to Crocker and Algina (1986), the following can account for restriction of range: incidental selection “made on the basis of some other variable that is correlated with the predictor test being validated” (p. 226); explicit selection “when the test being validated is used for selection purposes before its validity has been established” (p. 226), and natural attrition which occurs “whenever the subjects at the high or low end of the criterion continuum tend to leave the setting before criterion data can be collected” (p. 227).

Restriction of range produces conservative estimates of the correlations between the predictor and the criterion measures (Nunnally & Bernstein, 1994; Thorndike, 1949). In other words, restriction of the range of test scores or grades affects the estimated magnitude of the relationship between test scores or grades and the criterion variable; correlations are typically lower than expected and therefore, are not entirely accurate estimates of the true correlation between test scores or grades and subsequent performance for the total group of examinees. A demonstration by Givner and Hynes (1979) concluded that the more restricted the distribution of test scores, the more the correlation between the test score and the measure of achievement will be underestimated.

Criterion Unreliability

The aim of predictive validity studies is to establish the relationship between a predictor and criterion. According to Thorndike (1949), the magnitude of the relationship between the predictor and criterion depends on the precision with which these variables are measured. In practice, however, these two variables are assessed using empirical measures that tend to contain measurement errors (Muchinsky, 1996); these measures are considered “imperfect measures” of the constructs of interest (Kuncel, Campbell, Ones, 1998). Hence, the reliabilities of both the predictor and criterion have a tendency to be lower than their “true” reliabilities (Spearman, 1904).

Lack of prefect reliabilities of the predictor and criterion is viewed as problematic to validity results. “It limits the correlations that can be obtained between any predictor and criterion variable” (Thorndike, 1982, p. 197) and attenuates the magnitude of the relationship between the predictor and criterion or estimates of validity between these measures (Muchinsky, 1996). It is, therefore, necessary that validity coefficients obtained from the use of such predictors and criterion be corrected for predictor and criterion unreliability.

Some researchers have, however, suggested that only the criterion be corrected for unreliability because criterion unreliability attenuates the correlation between the observed scores on the predictor and the true scores on performance (Kuncel, Campbell, & Ones, 1998; Nunnally, 1978; Schimdt, Hunter, & Urry, 1976). Lee, Miller, and Graham (1982) stated that:

> Criterion unreliability should be corrected because it artificially obscures the optimal value of a selection instrument. Correcting for unreliability in the test is

> inappropriate since the objective is to estimate the predictor's actual (operational) value, not what its value would be if perfectly reliable (p. 637).

Adjustments to Compensate for Restriction of Range and Criterion Unreliability

When conducting validity studies it is desirable to adjust correlation coefficients to correct for error associated with restriction of range and criterion unreliability because "criterion unreliability and restriction of range attenuate a given validity coefficient, resulting in an underestimation of the true validity" (Raju & Brand, 2003, p. 52). These adjustments help determine the true relationship between the predictor and the criterion in the absence of range restriction and in the event that the criterion had perfect reliability.

Correcting validity coefficients for range restriction and criterion unreliability is recommended because it is assumed that the corrected validity coefficient is a superior estimate of the population correlation than the observed correlations (Thorndike, 1949; Gulliksen, 1950; Lord and Novick, 1968). Correcting for restriction of range and criterion unreliability are now widely accepted practices in educational and psychological testing and it is advised that both the corrected and uncorrected correlations be reported in validity studies (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999).

Procedures to correct for restriction of range and criterion unreliability are well established (See formulae in Appendix I). Validity coefficients are commonly corrected for restriction of range and criterion unreliability separately (Kelley, 1947; Gulliksen, 1950; Lord & Novick, 1968; Pearson, 1903; Spearman, 1904; Thorndike, 1949) or sequentially (Schmidt, Hunter, & Urry, 1976) in order to obtain a more accurate estimate of the true test validity.

Although the literature suggests that after correcting for restriction of range and criterion unreliability, standardized tests have a degree of relationship with subsequent performance such as future GPA or any measure of future academic performance, an important question that is constantly being asked is whether standardized tests are fair to the various groups of students (Linn, 1990). The various components of such questions are related to the issue of test bias (Wightman, 2003).

Test Bias and Admissions

The issue of test bias has received a lot of attention over the years. According to Cole (1972), possible bias in selection factors such as standardized tests is of great social and educational importance.

Although some researchers have made inferences to the fact that college admission tests help assess competence of examinees irrespective of their group membership (Jensen, 1980), others have disagreed with such statements. For example, some critics have stated that college bound students in the different subgroups have different standardized test scores, which usually favor Whites and men (Crouse & Trusheim, 1988). In the light of this comment, Crouse and Trusheim declared that colleges, especially selective colleges that insist that all examinees should have “equal” test scores, end up rejecting minorities and women who typically have lower scores than the majority or men; thus leading to the underrepresention of minorities and women in student populations. In the same vein, Nettles and Nettles (1999) mentioned that bias in standardized tests may create admission barriers to higher education for minorities and women. Based on such remarks, many egalitarians have argued that the use of the standardized tests to screen applicants for admission to colleges is biased against women and minorities.

A great deal of research has, therefore, been directed to the definition and identification of test bias (House & Keeley, 1993). According to Cleary (1968), "a test is biased for members of the subgroup of the population if, in prediction of a criterion for which the test was designed, consistent nonzero errors of predictions are made for members of the subgroup" (p. 115). Since then, other researchers have provided other definitions.

Shepard (1987) defined bias using one word, "invalidity". The implication drawn from Shepard's definition is that bias brings about systematic errors in the validity evidence gathered for a test. Shepard explained that that systematic error in a test distorts the meaning of the measurement for the members of a particular group. Jensen (1980) also defined bias as "systematic measurement error related to the use of a test with two or more specified populations" (p. 328). According to Linn (1984), bias is a systematic tendency for a test to over or under-estimate the true abilities of members of a group of examinees classified by demographic variables such as gender and ethnicity. Camilli and Shepard (1994) stated that test bias refers to "invalidity or systematic error in how a test measures for members of a particular group" (p. 8).

Accordingly, bias arises when test scores result in different meanings for the different identifiable subgroups. It is important to note, however, that test bias is defined in relation to groups of examinees and not the individual examinees in the group. Also, the fact that different groups have different average scores does not necessarily imply that the test is biased. Although large between-group difference may be indicative of test bias, score differences are not in themselves sufficient to establish the existence of test bias (Wightman, 2003).

To fully understand the underlying causes of group differences in test performance researchers have developed a number of theories to help the causes of test bias. Such theories include theories of performance differences in biology (Benbow & Lubinski, 1993; Halpern, 1992; Jensen, 1980), differences in brain lateralization (Halpern, Haviland, & Killian, 1998), and differences in social, psychological, and demographic factors (Austin, Clark, & Fitchett, 1971; Burton, Lewis, & Robertson, 1988; Borland, 1995; Young & Fisler, 2000).

As previously mentioned, test bias is a concept that is defined in terms of groups of examinees. Most often test bias is an issue in the study of gender and ethnic group differences (Camilli & Shepard, 1994, p. 8). To date, extensive research has been conducted on gender- and ethnicity-related bias on standardized tests.

Gender-Related Test Bias

For some time now, comparisons between the performance of men and women on school subjects and standardized tests have been a staple of educational and psychological researchers. The literature consistently reveals that men and women differ in their performance on various subjects and standardized tests with patterns of variation occurring at different points over time.

In a review of past research on gender differences in test performance, Wilder and Powell (1989) covered research that addressed undergraduate, graduate, and professional school admissions tests, validity studies, national studies, verbal ability tests, and quantitative ability tests. Specific testing programs discussed in the studies reviewed included the National Assessment of Educational Progress, National Longitudinal Study of 1972, High School and Beyond, and the SAT. The findings of the study revealed that women outperformed men on verbal ability and achievement tests while men outperformed women on mathematics tests. Although the study revealed that disparities existed between men and women, it also mentioned that these disparities were diminishing slowly over time.

A comprehensive examination of gender differences by Willingham and Cole (1997) also revealed gender differences across the different testing programs and in different subject areas. According to their findings, women tend to make better grades in school while men tend to make better scores on standardized tests.

Various researchers have gone on further to study gender differences in specific abilities and skills. Although some researchers have reported contradictory findings, the results regarding specific tests have generally shown that men tend do better in mathematics and science-related subjects and tests while women perform better on verbal subjects and tests (Azen, Bronner, & Gafni, 2002).

For example, a substantial body of evidence exists that suggests that from the beginning of secondary schooling, boys frequently outperform girls in mathematics at either age 9 or 13, but consistently though with relatively small differences at age 17 (Fennema & Carpenter, 1981, Meyer, 1989). Also, results from a meta-analysis conducted by Hyde, Fennema, and Lamon (1990) showed that while girls tend to do slightly better in mathematics compared to boys in the elementary and middle school years, by the high school and college years this difference changes and men tend to do much better than women.

Maccoby and Jacklin (1974) reported that women always have an advantage over men on verbal subjects and tests, and the difference became more pronounced over time. Hyde and Linn (1988) analyzed 165 studies covering 1,418,899 subjects that reported data on gender differences in verbal ability. Their results showed only a slight difference in women superiority with the difference being so small that it appeared to be nonexistent. Further, an examination of gender difference by age revealed countering results to the conclusions drawn by Maccoby and Jacklin (1974). That is, their analysis showed no striking changes in the magnitude of gender differences at different levels of ages.

Other researchers have studied gender differences in relation to test formats. Previous research has suggested that in several different subject areas, the average scores of men and women were nearly equal on essay portions while men had significantly higher average scores on multiple-choice sections of tests (Bridgeman & McHale, 1996).

Ethnicity-Related Test Bias

The general opinion held is that standardized test scores have less predictive power for Blacks and other minority populations (Fleming & Garcia, 1998; Shepard, 1987). Blacks typically score lower than Whites on vocabulary, reading, verbal, quantitative, analytical, and mathematics tests, as well as on tests that claim to measure scholastic aptitude and intelligence (Jencks & Phillips, 1998; Nettles & Nettles, 1999). For example, statistics provided by Nettles and Nettles indicated that average scores on both the verbal and quantitative components of the SAT, and on the verbal, quantitative, and analytical sections of the GRE are about 100 points lower for Blacks than for Whites.

Although researchers have indicted that the White-Black score gap appears to be narrowing, some studies have shown that the racial scoring gap still persists ("The Black-White Scoring Gap on SAT II," 2003) with Blacks typically scoring below Whites on most standardized tests for admission to medical, business, law, and other graduate programs. For example, Whites are five times, twelve times, and seven times more likely to score higher on the MCAT, Law School Admission Test (LSAT), and GMAT, respectively, than Blacks ("The Persisting Racial Scoring Gap," 2003). Bruschi and Anderson (1994) have also stated that there is still a large disparity between minority and majority students on science achievement tests. Less information is available regarding ethnic group performance differences on the different test formats such as essay tests in comparison to multiple-choice tests (Bridgeman & McHale, 1996).

Methods for Identifying Bias on Standardized Tests

The literature on identification of test bias generally focuses on two concepts: Differential item functioning (internal methods of test bias) and differential validity and differential prediction (external methods of test bias).

Differential Item Functioning

Differential item functioning (DIF) is employed to investigate test bias at the item level. Item bias is one of the potential threats to the validity of any test that occurs when a test item unfairly favors one group of examinees over another.

In DIF analysis, members of the different subgroups are matched on a measure of ability; typically total test scores. The probabilities of a correct response on a particular item for members of the subgroups are then compared (Linn, 1993). The assumption of DIF is that test takers with comparable abilities should perform equally on the individual test items regardless of their group membership. DIF is said to occur when a test item is found to be substantially harder for one group than for another group, after having controlled for overall differences in the ability levels of the groups.

Differential Validity and Differential Prediction

When tests are used for selection of students for admission and prediction of future performance, evidence of bias or lack of bias is generally sought in the relationships between test and criterion scores for the respective groups. When evidence is found by comparing the patterns of association between test scores and performance variables for the different groups, the term predictive bias may be used (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Predictive bias refers to the systematic error in predicting the criterion variable for particular groups of students (House & Keeley, 1993).

Two important concepts that are usually considered by researchers investigating prediction bias among examinees are differential validity and differential prediction. These two concepts are directly relevant to the issue of bias in selection decisions (Shepard, 1987).

Linn (1978, 1982) described differential validity as the differences in the magnitude of the correlation coefficients or validity coefficients for different groups of test-takers. Young (2001) also explained that differential validity refers to the situation where a test is predictive for all groups but to different degrees. In other words, the test lacks equivalent validities for the different groups of examinees. Questions concerning differential validity are, therefore, questions about whether the correlation between the predictors and the criterion are different for the different identifiable groups of examinees, and differential validity is examined by determining whether scores on a standardized test have the same predictive meaning or strength for members of different subgroups.

Differential prediction, on the other hand, refers to the situation where the derived prediction equations for the different subgroups of examinees are significantly different from group to group (Young, 2001). According to the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (1995), differential prediction exists if:

> Different algorithms are derived for the different groups and if the prediction leads to decisions regarding people from the individual groups that are systematically different from those decisions obtained from the algorithm based on the pooled groups (p. 12).

Questions about differential prediction, therefore, are questions related to whether the prediction models obtained for the different subgroups of examinees are different. Such questions are generally approached by comparing regression systems for equality of regression slopes and equality of regression intercepts in the respective prediction equations across the groups of interest (Linn, 1978, 1982; Shepard, 1987).

Both differential validity and differential prediction are important in the validation of the scores on a test. However, of the two issues, differential prediction is the more crucial because differences in prediction have more relevance when considering fairness in selection than do differential validities (Linn, 1982). Nevertheless, research on both differential validity and differential prediction are important because they consistently show evidence of differences in the prediction equations across subgroups of examinees.

Previous Studies on Differential Validity and Differential Prediction of Standardized Tests

Thus far, a number of studies has been conducted specifically on differential validity or differential prediction, as well as on the combination of differential validity and differential prediction of admissions tests such as the SAT and GRE.

Breland (1979) conducted a comprehensive review of a number of studies on differential validity and differential prediction from 1964 to 1974. Predictors used in these studies included SAT mathematics, SAT verbal, and high school rank; the criterion measure employed was freshman grades. Breland's findings suggested that in terms of differential validity, the median values of the predictors for women were generally equal to or higher than that for men. However, validity coefficients across the different ethnic groups showed no discernible pattern. In terms of differential prediction, Breland concluded that college performance of minority students was consistently overpredicted and the degree of overprediction was more pronounced for Blacks than for other minority groups.

Other studies on differential validity and differential prediction have also shown that, in general, admission test scores tend to overpredict future grades for minority students. That is, minority students, especially Blacks, tend to earn lower grades than were predicted from their test scores (Cleary, 1968; Ramist, 1984; Young, 1991, 1994).

Thomas (1972) compared prediction equations for men and women at 10 colleges. The results obtained indicated that women's GPA was underpredicted when a similar prediction equation was used to predict freshman GPA for men and women. Implying that, women achieved higher GPAs than predicted.

Young (2001) reviewed published studies on differential validity and prediction dating back to 1974. His review covered 49 studies on differential validity and differential prediction across gender and ethnicity. Young reported that:

> in general, multiple correlations computed from samples of Black or Hispanic students are somewhat lower than for Asians and Whites (p. 12) and with a few exceptions. The findings consistently point to an overprediction of the grades of Black and Hispanic students' grades (p. 15).

Studies related specifically to the GRE by House (1998) found out that while the GRE was generally predictive of graduate performance, in a number of cases it underpredicted the achievement of women and overpredicted the achievement of men.

Researchers have tried to offer an explanation for such results. For example, Burton and Ramist (2001) asserted that test scores do not necessarily underpredict women future academic performance but rather, the actual grades obtained by women are higher than predicted because women tend to enroll in less stringently graded courses. Some studies that have adjusted prediction equations for differences in college grading patterns have shown that the appearance of bias is indeed reduced or completely eliminated (Elliot & Strenta, 1988).

Medical School Admissions

Medical school admissions committees are faced with the perennial problem of selecting candidates who are properly prepared and highly motivated for the medical program (Hanlon, 1964), have the inclination to understand and improve the human condition (Elam, Seaver, Berres, & Brandt, 2000), and will successfully complete the medical training, become good physicians, and serve societal needs for health care delivery (Gough, Hall, & Harris, 1963; Manning, Willingham, & Breland, 1977). Thus, selection processes of medical students aim at two goals: short-term goals, which concerns students' progress through the curriculum, and long-term goals, that is, students' ability to provide clinical care and advancing health care (McGaghie, 2002b). The decision to admit an individual to medical school is, therefore, a very important one; it is tantamount to a decision to grant them medical training and a license to practice medicine (Johnson, 1983; McGaghie, 1987).

Admissions committees use different screening practices to select students. In some medical schools a subset committee is given the primary task of previewing admission files. Other admissions committee members then rely on the recommendations of the subset committee to make their decisions (Elam, Stratton, & Lieber, 2002). In other settings, all admissions committee members review the admission files to achieve a comparable level of familiarity with the applicants (Elam & Johnson, 1997).

Medical school admissions committees use selection criteria in their decision-making. The criteria for admission to medical school have come a long way since the early nineteenth century when medical school requirements were not standardized (Ludmeyer, 1999) with only a few colleges having university affiliations. As a consequence, new doctors appeared on the scene every eight months, charged with the responsibility, but not the ability, to tend to the nation's medical needs (Iserson, 1997).

Medical school admission requirements were finally standardized in 1894 (Iserson, 1997). The 1894 requirements consisted of a handwritten English composition of not less than 200 words (the said composition to include construction, punctuation, and spelling); arithmetic including fundamental rules, decimals fractions, ratio and proportion, algebra with quadratics, and a year's study of Latin. Medical schools also required a high school diploma or its equivalent as demonstrated by examination.

By the mid 1900s the criteria for medical school expanded further to include personal interviews, letters of recommendation from undergraduate faculty, undergraduate GPA, and MCAT test scores. In a 1963 appraisal of admission procedures, Gough, Hall, and Harris (1963) asserted that three lines of evidence were utilized to evaluate applicants for medical training. These were premedical scholastic achievement, MCAT, and appraisals drawn from a personal interview.

The current medical school criteria includes biographic information, postsecondary experiences, letters of evaluation written by faculty members, extracurricular community activities (Elam, Seaver, Berres, & Brandt, 2002), a year of biology, general chemistry, organic chemistry, physics, English composition, and English literature or a communication course (Elam, Taylor, & Strother, 1996). In addition, most medical school admissions officers expect that applicants will complete their baccalaureate degrees prior to matriculation into the medicine program.

Most medical schools also require or suggest an interview for the most qualified candidates as part of their selection procedure (Gottheil & Michael, 1957). The interview is used to further discuss applicants' qualifications, such as their motivation, interests, leadership qualities, ability to deal with pressure, work with and care about others, academic performance, and professional goals (Elam, Burke, Wiggs, & Speck, 1998;

Johnson, & Edwards, 1991; Nowacek, Bailey, & Sturgill, 1996; Powis, McManus, & Cleve-Hogg, 1992).

Almost all medical schools require students making applications to take the MCAT as part of the admission requirements. Medical schools that do not require the MCAT even view students who submit their MCAT scores as having a considerable advantage over those who do not submit their test scores (Iserson, 1997).

Although all admissions committees consider academic and nonacademic characteristics of students as important to the admissions process, admissions practices and procedures vary for the various schools. A survey conducted by Mitchell, Haynes, and Koenig (1994) to examine admissions practices and the use of MCAT data in student selection showed that almost all admissions committees gave the most weight to MCAT scores because the MCAT helps identify the most academically capable individuals in their applicant pool. The study also reported that some admissions committees rely heavily on the individual MCAT subtest scores and use the individual subtest scores to predict students' academic performance in the medical school.

GPA and MCAT Scores as Predictors of Students' Performance in Medical School

Various studies show that MCAT test scores and GPA are the most important criteria for medical admissions (Olmstead & Sheffrin, 1981). Moss (1938) demonstrated the effectiveness of GPA and the MCAT in medical school admissions. Based on his analysis he concluded that MCAT and GPA measured some of the traits necessary for success in medical school, and also that MCAT scores predicted first year medical school grades more accurately than GPA. Friedman and Bakewell (1980) found out that the prediction of first-year medical student performance was increased by the MCAT on the order of 30 percent over that which is possible with all other admission variables.

McGuire (1980) employed a slightly different criterion in his study. He looked at how well MCAT scores predicted medical school freshman class standing. His results showed that MCAT was the single best predictor of freshman class standing. Colliver, Verhulst, and Williams (1989) also established the validity of MCAT and undergraduate GPA as the main predictors for students' performance both in clinical and basic science courses.

Mitchell (1991) estimated the predictive validity of the 1991 MCAT using undergraduate GPA, institutional selectivity, and examinees' scores on Verbal Reasoning, Physical Sciences, and Biological Sciences items. Four criterion measures were considered in the study. The median multiple correlation between the combination of the predictors and the criterion measures, year one, year two, year three, and year four medical school GPA, were .69, .62, .60, and .54, respectively. The results indicated that the MCAT was useful for predicting performance in medical school; for admission officials for identifying applicants most likely to succeed in medical school; for assessing strengths and weaknesses in knowledge of entry-level science content, science problem solving and critical analytical thinking; and for interpreting applicants' transcripts and letters of evaluation from the applicants' undergraduate institutions.

In a study to address the usefulness of undergraduate GPA and MCAT scores in predicting first-year medical school GPA, Mitchell, Haynes, and Koenig (1994) reported that the median corrected multiple correlation between medical school year-one GPA and undergraduate GPA, medical school year-one GPA and MCAT, and medical school year-one GPA and the combination of undergraduate GPA and MCAT were .53, .66, and .73, respectively.

Wiley and Koenig (1996) evaluated the predictive validity of the MCAT for the first two years of medical school. The analysis included entrants to 16 medical schools in the U.S. Preadmission data for each student included GPA, MCAT composite scores, and an undergraduate institutional selectivity index. The results pointed towards a significant relationship between MCAT scores and subsequent medical school performances.
The median correlations between cumulative year-one and year-two medical school GPA and MCAT scores were reported in the range of .62 to .67 for the 16 medical schools. On the other hand, median correlations between cumulative year-one and year-two medical school GPA and undergraduate GPA ranged from .54 to .58. When the two predictors were combined median correlations between the criterion measure and the predictors ranged from .70 to .76. Adding institutional selectivity to regression models that already included GPA and MCAT scores did not substantially increase these predictive values.

After correcting for restriction in range, Julian and Lockwood (2000) found out that the median coefficients between undergraduate GPA and cumulative year-one and year-two medical school GPA, between MCAT and cumulative year-one and year-two medical school GPA, and between the combination of MCAT and undergraduate GPA and cumulative year-one and year-two medical school GPA were .54, .59, and .71, respectively.

Koenig, Huff, and Julian (2002) studied the predictive validity of GPA and MCAT scores for the first three years of medical school for 1992 and 1993 entrants to 14 medical schools. The corrected median correlations for year-one, year-two, and year-three medical GPA and the combination of undergraduate GPA and MCAT scores were, .70, .64, and .53, respectively.

Overall, the results on the predictive validity of the MCAT show that MCAT scores make a significant contribution to the prediction of success in medical school. However, these results were obtained from studies that used MCAT scores as a composite. Thus far, very few studies have been done on the predictive validity of MCAT subtest scores. Although it is believed that some medical schools may have already conducted such studies, a library search failed to identify published or disseminated studies on the predictive validity of MCAT subtest scores specifically for the current version of the MCAT.

Predictive Validity of Subtest Scores for Older Versions of the MCAT

Findings of the few studies that investigated the relationship between subtest scores of the older versions of the MCAT and students' performance in medical school showed that although significant relationships existed, their magnitude became much weaker as a student progresses through medical school (Mitchell, 1990).

Hill (1959) investigated the predictive validity of the MCAT subtests, Quantitative Ability, Verbal Ability, Science, and Understanding Modern Society. The study included 1,000 students from a New York medical college between 1950 and 1957. First-year medical school GPA was used as a criterion variable. Hill's study demonstrated a relationship between first-year GPA and three of the MCAT subtest scores, namely, Quantitative Ability, Science, and Understanding Modern Society. The correlation between MCAT Verbal and first-year medical school GPA, on the other hand, was not statistically significant. Hill, however, did not provide inferences regarding such relationships.

Cullen, Peckham, and Schwarz (1980) also studied the predictive validity of the subtest scores on the 1977 MCAT for first-quarter medical school GPA. The results of their study indicated that the Science subtest was the most useful in predicting students' performance, followed by the quantitative subtest, and then the verbal subtest.

Ramos, Croen, and Haddow (1986) also found out that the MCAT subtests do measure students' abilities that are relevant to the study of medicine. Their findings showed that the science subtest correlated very highly with first- and second-year medical school GPA.

Again, in a study to determine the relative importance of scores on the Science, Verbal and Quantitative subtests of the 1977 MCAT in predicting performance during medical school, results obtained by Glaser, Hojat, Veloski, Blacklow, and Goepp (1992) suggested that the Science subtest had a greater influence on students' performance in the early years of medical school.

However, the results from studies conducted by Jones and Thomae-Forgues (1984) and Jones and Vanyur (1985) counteracted these findings. Their studies suggested that the Verbal subtest was a better predictor of students' performance in medical school than the Science subtest.

Differential Validity and Differential Prediction of MCAT Scores

According to the AAMC (1986), performance differences are evident on the MCAT for population subgroups differentiated by sex and ethnic status. Other empirical studies on gender and ethnic performance differences on the MCAT have also reported similar findings.

A study by Halpern, Haviland, and Killian (1998) on sex differences on the MCAT scores reported that men generally scored higher than women on tests of scientific knowledge. Jones and Mitchell (1986) examined ethnic bias in relation to performance variance in medical school. The sample in the study consisted of all Black and White entrants to American medical schools in 1978 and 1979. The results indicated that differences existed on mean MCAT scores for Black and White students. Mean MCAT scores on the subtests (Biology, chemistry, physics, science, verbal/reading skills analysis, and quantitative skills analysis) for White matriculants ranged from 9.4 to 9.8, compared to 6.3 to 7.1 for Blacks. The mean subtest differences between Black and White enrollees ranged from 2.4 to 3.3.

In regards to differential validity, Jones and Vanyur (1985) found out that the correlations between MCAT scores and undergraduate GPA with first- and second-year GPAs tend to be stronger for women than for men. Studying ethnicity-related differential validity of the MCAT, Johnson, Lloyd, Jones, and Anderson (1986) examined the validity of the MCAT, UGPA, and "competitiveness" of 30 selected undergraduate colleges in predicting the performance of students at a predominantly Black medical school. The performance measure consisted of course grades in all four years of medical school. The validity coefficients estimated for the students at the Black college were similar to those revealed in earlier studies conducted at predominantly White schools.

Various studies on differential prediction of MCAT scores by gender and ethnicity have, however, reported conflicting results. Jones and Vanyur (1985) failed to identify the presence of significant gender-related bias in the MCAT for the first two years of medical school at two midwestern state-supported medical schools. Further, the study failed to identify widespread bias in predicting women GPA from MCAT scores and undergraduate GPA.

In a validity study to identify bias in the use of GPA and the MCAT in the selection of medical students, Vancover, Reinhart, Solomon, and Haf (1990) also found out that GPA and the MCAT were indeed valid and equally predictive for minority and majority groups. Also, to examine the predictive validity of MCAT scores across gender and ethnic groups, Koenig, Sireci, and Wiley (1998) studied 1992 entrants to 14 American medical schools. Criterion measures included students' cumulative GPA in the first two years of medical school. Differential predictive validity was examined by comparing prediction errors across gender and ethnicity. The patterns of prediction errors showed that on the average there were no evident differences between men and women. However, performances of Whites tended to be underpredicted, although the magnitude of difference in prediction was very small. On the other hand, performances of Blacks, Asians, and Hispanics tended to be overpredicted with significant findings for Asians and Hispanics.

Summary

The admission of poorly qualified students to any program misuses resources of schools, faculty, and students (Kuncel, Hezlett, Ones, 2001). Hence, effective selection of students into any field of study is of critical importance. In order to identify well-qualified students for their curriculum, admissions committees employ a set of admission requirements, which include both cognitive and non-cognitive factors. A historical overview of college admission requirements is presented at the beginning of this chapter.

Although non-cognitive admission factors such as interviews, letters of recommendation, and so forth, play a role in the admissions process, research indicates that admissions committees rely more heavily on indicators of students' cognitive ability such as standardized test scores and previous grades in their decision-making; some of the reasons being that these criteria are observable and easily quantifiable. Given their widespread use in admissions, it is essential that the validity of standardized test scores and previous grades be investigated to ensure the accuracy of admission decisions (Young, 2001).

Predictive validity is one aspect of validation used to describe how well scores on selection criteria predict future performance. Jensen (1980) referred to this kind of validity as "the most important, defensible, and convincing type of validation in the practical use of psychological tests" (p. 298). In predictive validity studies test scores and grades are used as predictors and students' future performance, the criterion measure. In evaluating the validity of test scores and GPA, it is important to note that the validity coefficients, obtained as a measure of the relationship between the predictor(s) and the criterion measure, are typically affected by factors such as restriction of range and criterion unreliability.

Restriction of range and criterion unreliability both attenuate validity coefficients, resulting in an underestimation of the true relationship between the predictors and the criterion. Researchers have proposed a number of procedures to correct for restriction of range as well as for criterion unreliability. The literature consistently indicates that, after correcting for restriction of range and criterion unreliability, test scores and previous grades better predict students' performance across various testing programs.

In addition to predictive validity, issues of test bias are also considered to be important in validation research because "they have relevance for the issues of test bias and fair test use" (Young, 2001, p. 4). Differential validity and differential prediction are two concepts of test bias. Differential validity deals with whether a test is equally predictive across identifiable subgroups, and differential prediction involves determining whether a similar prediction equation could be used to equally predict subsequent performance for all subgroups.

Generalizations made from studies on differential validity suggest that the validities of test scores and previous grades are comparable for both majority and minority ethnic groups of students but are at times higher for women than for men. In respect to differential prediction, however, when a single prediction equation was used for the different ethnic and gender subgroups, future performance for minority students was overpredicted and future performance for women was underpredicted.

Studies on the predictive validity of the MCAT and undergraduate GPA, which have been identified as the most important criteria by which medical students are selected for admission, for instance, showed that these two factors do predict students' performance on the medical program. Conflicting results have, however, been reported with regards to the differential validity and differential prediction of the MCAT. Whilst some researchers have concluded that predictive validities of the MCAT did not differ across subgroups, others have indicated that correlations between the MCAT scores and students' future performance were typically different for the men and women. Also, with respect to differential prediction of the MCAT, a number of empirical studies have concluded that bias does not exist across gender and ethnicity. At the same time other findings have identified the existence of ethnicity-related bias for the MCAT.

Medical school admissions committees are continually interested in empirical studies that examine the relationship between medical school predictors and students' subsequent performance in medical school. To date, most of such studies used MCAT composite scores and aggregate student data in their data analysis. Although these studies continue to provide relevant information for admissions, researchers have suggested that there is also a need to examine the predictive validity of the individual subtests of the MCAT since some admissions committees place greater emphasis on applicants' subtest scores.

In addition, studies on differential validity and differential prediction of MCAT subtests across subgroups such as gender and ethnicity subgroups are important topics worthy of investigation. Such information would be useful for medical school admissions committees, especially those who consider the individual MCAT subtest scores in admissions, to ensure that the subtests do predict medical school performance and that the scores on the subtests are not used in a manner that is consistently unfair to certain groups of examinees.

Chapter Three: Methodology

This chapter provides a detailed description of the methodology used in this study. The discussion is divided into two sections. Section one presents information on instrumentation and the research design employed in the study. In section two, data analysis procedures utilized in the study are discussed.

Instrumentation

The instrument used in this study is the MCAT. The MCAT is administered to aspiring medical students twice a year under standardized, fair, equal, and secure testing conditions. The test contains four subtests: Biological Sciences, Verbal Reasoning, Physical Sciences, and Writing Sample.

Item Format and Test Specifications

The Biological Sciences subtest covers topics on biology and organic chemistry while the Physical Sciences subtest includes questions on physics and general chemistry. The Verbal Reasoning test consists of questions from the areas of humanities, social sciences, and natural sciences. Skills measured on the Verbal test include comprehension (24 questions), evaluation (13 questions), application (17 questions), and incorporation of new information (11 questions).

The Biological Sciences, Verbal Reasoning, and Physical Sciences subtests of the MCAT are composed of multiple-choice items. Response options range from A to D for each item and responses are scored dichotomously as either 1 (*correct*) or 0 (*incorrect*), with no penalty for incorrect responses.

The Biological Sciences and Physical Sciences sections each contain 77 items with about 11 passage-based sets and 4 to 8 items per passage. The two subtests also contain about 14 discrete items. A fixed number of items are experimental, and therefore not scored. The Verbal Reasoning section, on the other hand, contains nine passage-based sets with about 5 to 10 items per passage. As with Biological Sciences and Physical Sciences, a fixed number of items are experimental.

The Writing Sample subtest consists of essay items with scores ranging from J to T; or 1 to 11 when assigned numeric values. Areas assessed by this subtest are as follows: developing a central idea; synthesizing concepts and ideas; presenting ideas cohesively and logically and writing clearly; following accepted practices of grammar, syntax, and punctuation, consistent with timed, first-draft composition.

Equipercentile equating is used to adjust scores on different forms of the MCAT to compensate for differences in difficulty among the test forms and to maintain a single score scale for all forms of the MCAT. Raw scores on each MCAT subtest are converted to scaled scores by way of a designed formula. The scaled scores are reported to the medical schools as follows: Biological Sciences, Physical Sciences, and Verbal Reasoning are reported on a scale of 1 to 15. The Writing Sample subtest is converted into a scale of 1-11. Each MCAT administration consists of new items, and previously administered (anchor) items, which make possible equating back to previous administrations. Table 1 shows the number of items, maximum points, and time limits for each subtest.

Table 1

Number of Items, Maximum Points, and Time Limit for MCAT Subtests

Subtest	Number of Items	Maximum Points	Time Limit
Biological Sciences	77	15	100
Physical Sciences	77	15	100
Verbal Reasoning	65	15	85
Writing Sample	2	11	60

Research Design

The criterion measure used in this study is first-year medical school GPA. Predictor sets include:

1. Biological Sciences scores alone.
2. Verbal Reasoning scores alone.
3. Physical Sciences scores alone.
4. Writing Sample scores alone.
5. Undergraduate non-science GPA alone.
6. Undergraduate science GPA alone.
7. The combination of Biological Sciences, Verbal Reasoning, Physical Sciences, and Writing Sample subtest scores.
8. The combination of undergraduate non-science and science GPA.
9. The combination of Biological Sciences, Verbal Reasoning, Physical Sciences, Writing Sample subtest scores, and undergraduate non-science and science GPA.

10. MCAT total/composite scores, which is derived as a sum of the scores on the Biological Sciences, Verbal Reasoning, Physical Sciences, and Writing Sample subtests.
11. The combination of MCAT total/composite scores, and undergraduate non-science and science GPA.
12. A subset/block of Biological Sciences, Verbal Reasoning, and Physical Sciences subtest scores with a subset/block of the Writing Sample subtest scores.

Using these predictor sets and the criterion measure, the study first evaluated the predictive validities of the MCAT subtests scores and undergraduate GPA, individually and in combination. Differential validity and differential prediction of MCAT subtest scores, individually and collectively, were then examined across examinees' gender and ethnicity. Specific gender and ethnicity comparisons included men versus women, Whites versus Blacks, Whites versus Asians, and Whites versus Hispanics.

Data Source

The data employed in the study were obtained from the MCAT Predictive Validity Research (PVR) database with the approval of the MCAT section of the Association of American Medical Colleges (See Appendix A). Also, exempted permission to conduct this study was obtained from the Ohio University Institutional Review Board (Appendix B) based on the fact that the data did not contain codes or identifiers that may be linked to any of the participants of the study.

The data used in the study consist of preadmission information collected for two cohorts, 1992 and 1993 matriculants to 14 medical schools in the U.S. The research division selected the 14 medical schools in order to obtain a geographically, ethnically, and administratively representative sample of the 125 medical schools in the U.S. The preadmission information included undergraduate non-science and science GPAs, and Biological Sciences, Physical Sciences, Verbal Reasoning, and Writing Sample subtest scores for each student. First-year medical school GPAs were also provided for each student by the participating schools.

Sample

Data obtained from the MCAT research division consisted of 3,187 medical students selected from the 1992 and 1993 entering classes of 14 medical schools. The number of students in a school's cohort ranges from 64 to 148 for the 1992 cohort and 0 to 279 for the 1993 cohort, with a median of 106 for both cohorts.

The sample comprised of 58.8 % men and 41.2% women (Table 2). Ethnic group information indicated that 60.7 % are Whites, 13.4 % are Asians, 10.6 % are Blacks, 6.0 % are Hispanics, and 1 % percent are American Indians (Table 3). Information on ethnic identity was unavailable for 8.3 % of the total study sample.

Table 2

Frequency Distribution of Gender of Students

Gender	Frequency	Percent
Men	1,875	58.8
Women	1,312	41.2
Total	3,187	100.0

Table 3

Frequency Distribution of Ethnicity of Students

Ethnicity	Frequency	Percent
Whites	1,936	60.7
Blacks	338	10.6
Asians	427	13.4
Hispanics	191	6.0
American Indians	31	1.0
Not Indicated	264	8.3
Total	3,187	100.0

Research Questions

The study addressed five research questions.

1. How well do undergraduate GPA and MCAT subtest scores, both individually and collectively, predict first-year medical school GPA?
2. How well do MCAT total/composite scores predict first-year medical school GPA?
3. How well do subsets of MCAT subtest scores predict first-year medical school GPA?
4. Are the predictive strengths of MCAT subtest scores, both individually and collectively, consistent across gender?
5. Are the predictive strengths of MCAT subtest scores, both individually and collectively, consistent across ethnicity?

Data Collection Procedures

Preadmission data such as undergraduate GPA and Biological Sciences, Physical Sciences, Verbal Reasoning, and Writing Sample scores, were collected for 1992 and 1993 medical school entrants to 14 medical schools as part of the their medical school application process. First-year medical school grades for these students were also provided. Data were collected from these two cohorts of 14 medical schools because the schools agreed to participate in the MCAT Predictive Validity Research.

Data Analysis Procedures

Since each medical school has its own grading procedure, the criterion measure, first-year medical GPA, is school dependent. Data analyses were therefore carried out separately for the 14 medical schools that participated in the study. The alpha level for each statistical procedure was set at .05.

Data analyses were completed in three phases to help answer the research questions posed in the study.

Phase 1: Descriptive Statistics

Student counts and means and standard deviations for the Biological Sciences, Physical Sciences, Verbal Reasoning, and Writing Sample subtest scores, and for undergraduate non-science and science GPAs were computed for students in the individual medical schools and over all medical schools. On the other hand, criterion means and standard deviations were only calculated for the individual medical school. Means and standard deviations for the predictors and the criterion were also computed by gender and by ethnicity.

Phase 2: Research Questions 1 to 3 - How well do undergraduate GPA and MCAT subtest scores, both individually and collectively (question 1), MCAT total/composite scores (question 2), and subsets of MCAT subtest scores (question 3) predict performance in the first year of medical school?

To help answer the questions related to the predictive validities of the MCAT subtest scores and undergraduate GPA, simple linear and multiple regression analyses (these statistical procedures are described in detail subsequently) were performed. The data were first screened to assess whether they fit the assumptions of regression analysis (Mertler & Vannatta, 2002). Scores on the Biological Sciences, Verbal Reasoning, Physical Sciences, and Writing Sample subtests, and undergraduate non-science and science GPAs were entered as indicated by the predictor sets into regression models to predict first-year medical school GPA. Zero-order and multiple correlations between the predictor sets and first-year medical school GPA were then examined and reported for each medical school and across the 14 medical schools.

Phase 3: Research Questions 4 to 6 - Are the predictive strengths of MCAT subtest scores, both individually and collectively, consistent across gender (question 4) and ethnicity (question 5)?

Differential validity and differential prediction of the Biological Sciences, Physical Sciences, Verbal Reasoning, and Writing Sample subtests, as well as the combination of these subtests, were assessed across examinees' gender and ethnicity.

Differential validity was assessed by transforming the validity coefficients obtained for the different gender and ethnicity subgroups into Fisher's z values; the z values were then tested for significant differences (Edwards, 1962), by way of the following equation:

$$Z = \frac{z_1 - z_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$$

where z_1 and z_2 are the correlations expressed as Fisher z values for groups 1 and 2, respectively, and N_1 and N_2 are the sample sizes for groups 1 and 2, respectively.

Differential prediction of the MCAT subtest scores was evaluated using a procedure proposed by Gulliksen and Wilks (1950) by which standard errors of estimate, regression slopes, and regression intercepts obtained for different subgroups are compared for significant difference; this is the accepted approach for examining differential prediction (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

In addition, common regression equations were fitted for all students within a school and then patterns of prediction errors (the difference between the actual and predicted first-year medical school GPAs) for the various subgroups were examined (Anastasi, 1988; Young, 2001). According to Young, overprediction occurs when “the residuals from a prediction equation based on a pooled sample are generally negative for a specific group” and conversely, underprediction occurs when “the residuals are generally positive.” Simply speaking, overprediction occurs when a subgroup does not perform as well as predicted. That is, their predicted performance is above their actual performance. Conversely, underprediction occurs when a subgroup performs better than predicted or their predicted performance is below their actual performance.

Statistical Methods for Detecting Predictive Validity of Admission Tests

Over the decades researchers have advanced the knowledge of assessing the predictive validity of admission tests. Mitchell (1990) identified four statistical methods “commonly used in local and national-level predictive validity studies” (p. 150). These are discriminant function analysis, structural equations modeling, correlation analysis, and regression analysis.

The intent of discriminant function analysis, often known as discriminant analysis or simply DA, is to classify subjects into groups or predict group membership from a set of predictors (Kerlinger & Pedhazur, 1973; Mertler & Vannatta, 2002; Pedhazur, 1997; Tabachnick & Fidell, 2001). On the other hand, Pedhazur explained that structural equation modeling is a confirmatory procedure used to study patterns of causation among variables. Tabachnick and Fidell also stated that “the technique evaluates whether the model provides a reasonable fit to the data and the contribution of each of the independent variables to the dependent variables” (p. 26).

“The vast majority of investigations are correlational; some report simple correlations for pairs of predictor and criteria” (Mitchell, 1990, p. 150). According to Pedhazur and Schmelkin (1991), correlation is the basic approach used to study the relationship(s) between the predictor(s) and the criterion. The index of such a relationship is the correlation coefficient, which is at times referred to as the validity coefficient and represented as r_{xy}. The disadvantage of using correlation analysis in predictive validity research is that the correlation coefficient provides researchers with only a general sense of the magnitude and direction of the relationship between the predictor and the criterion.

For instance, a positive correlation implies that as the values on one variable increases, the values on the other variable also increases in the same direction.

Many predictive validity studies employ regression as the statistical procedure in their data analysis (Mitchell, 1990). “Regression analysis is a method of analyzing variability of a dependent variable or a criterion by resorting to information on one or more independent variables or predictors” (Pedhazur & Schmelkin, 1991, p. 371). According to Crocker and Algina (1986), regression analysis is “the appropriate” statistical procedure for assessing the effectiveness of predictors in validation studies. Also, Pedhazur and Schmelkin stated that:

> To be useful (e.g. for selection), a predictive system should enable one to make more specific predictions about expected status on the criterion, given status on the predictor. One of the most useful means of accomplishing this is through the use of a prediction equation obtained in a regression analysis (p. 38-39).

Regression Analysis

The objective of regression is to predict a dependent variable from an independent variable. The primary purpose of regression analysis is to develop an equation that explains the relationship between the dependent variable and the independent variable and also to predict the values on the dependent variable for a given population (Mertler & Vannatta, 2002). Regression equations are used to examine “how differences in one variable relate to differences in another and allow us to predict a person’s score on one variable from knowledge of that person’s score on another variable” (Howell, 1995, p. 167).

The basic idea of linear regression is that the relationship between the independent variable and the dependent variable can be explained with a straight (best-fitting to the data points) line with the equation:

$y = ax + c$

where a is the slope and c, the intercept. In regression analysis, the mathematical equation is often presented as follows:

$Y_i = B_o + B_i X_i + e_i \quad i = 1, 2, 3,\ldots, n$

Y_i is the observed value of the dependent variable/criterion. B_o symbolizes the intercept constant; the value of the dependent variable when the independent variable equals 0. B_i indicates the slope for the independent variable (X_i); the amount of change in the independent variable with a unit change in the dependent variable. Finally, e_i stands for prediction error or residual; the difference between observed value and predicted value of the dependent variable. The main aim of regression procedures is to minimize the amount of prediction errors so as to maximize the linear relationship between the independent variable and the dependent variable (Kerlinger & Pedhazur, 1973).

In the case where there are multiple predictors, multiple regression is employed. “In multiple regression we are interested in predicting a dependent variable from a set of predictors” (Stevens, 1996, p. 64). Kerlinger and Pedhazur (1973) also mentioned that, “multiple regression is a method of analyzing the collective and separate contributions of two or more independent variables, X_i, to the variation of a dependent variable, Y” (p. 3). According to Tatsuoka (1969), multiple regression is a multivariate procedure used for validating tests for admission, employment, and promotion. The task of multiple regression is to help us explain the variance of a dependent variable by estimating the contributions of two or more independent variables to this variance.

Issues Peculiar to Multiple Regression

Mertler and Vannatta (2002) identified three issues peculiar to multiple regression: Multiple Correlation, Multicollinearity, and Model Selection.

Multiple correlation, which is symbolized by R, is an index of the relationship between the observed and predicted values on the dependent variable. The squared multiple correlation, R^2, represents the proportion of variance in the dependent variable accounted for by the independent variables (Pedhazur, 1997).

Multicollinearity occurs when the predictor variables in a study tend to have moderate to high intercorrelations among each other (Mertler & Vannatta, 2002; Stevens, 1996). To detect multicollinearity, there is the need to first examine the simple correlations among the predictors. In addition, tolerance and variance inflation factor (VIF) values for the predictors must be examined. Tolerance is a measure of collinearity among the independent variables with possible values ranging from 0 to 1; a value close to 0 being an indication of multicollinearity. VIF is the inverse of tolerance and indicates whether there is a strong linear association between a given predictor and all the other predictors. Values greater than 10 are generally a cause for concern (Stevens, 1996).

Tabachnick and Fidell (2001) outlined three methods for model selection in regression analysis.

1. Standard Multiple Regression

With this type of regression modeling (often referred to as the full model) all of the independent variables are entered at the same time into the model and the amount of variance explained by each independent variable is assessed.

2. Sequential Multiple Regression

This method is at times known as hierarchical regression. Instead of just entering all the independent variables into the model as in the case of standard multiple regression, the independent variables are entered into the equation in a particular order or sequence to examine the effect of each variable.

3. Stepwise Multiple Regression

This method is often referred to as statistical multiple regression. There are three forms of stepwise regression:

Forward selection where the variable with the highest correlation is entered into the equation first and its contribution to R^2 is assessed before the second variable is entered. The second variable that adds the most significant amount of unique variance to the model after the first variable is accounted for is entered, and then the change in R^2 is examined. Other variables are entered following this procedure until the last variable added ceases to contribute significantly to the model.

Stepwise selection where after each variable is introduced into the model its significance in the model is assessed until the best fitting model is obtained.

Backward deletion where all the independent variables are entered into the equation and then a partial F test is conducted for each variable as if it were the last variable to enter the equation. At each stage, the change in R^2 is examined.

Variations of these models have been applied in predictive validity studies. For the purpose of this study, stepwise regression was used because according to Tabachnick and Fidell (2001) stepwise regression is typically used to develop a subset of predictors that is useful in predicting the criterion. The stepwise model also helps determine whether a particular order of entry existed for the predictors used in the study. Decisions about the entry of the predictors into the equation are based on the amount of correlation between each predictor and the criterion. The predictor which has the highest correlation with the criterion or that explains the most unique variance will be entered first into the model. Next, the predictor that appears to explain the most unique variance or contributes more strongly to the squared multiple correlation after the contribution of the first predictor is accounted for will be entered into the equation. Predictors will continually be added in this manner until no other predictor contributes unique variance to the regression model.

Assumptions of Regression Analysis

According to Fox (1997), the assumptions of regression analysis in research where the independent variables or predictors are not fixed are as follows:

1. The scores on the independent variable are random.
2. The expected value of the dependent variable is a linear function of the independent variable.
3. The scores on the independent and dependent variables follow a joint normal distribution.
4. The independent variable and error values are assumed to be independent in the population from which the sample is drawn.
5. The error values follow a normal distribution.

Testing the Assumptions of Regression

There are essentially two approaches to testing multiple regression assumptions (Tabachnick & Fidell, 2001). The first approach involves screening the data for linearity and normality using scatterplots and histograms. Normality can also be assessed using Kolmogorov-Smirnov statistics, as well as an examination of skewness and kurtosis. The second approach involves the examination of the errors or residuals by way of histograms of the standardized or studentized residuals (Stevens, 1996).

Testing for Differential Prediction

This approach involves a test of equality of standard errors of estimate, followed by a test of equality of regression slopes, and then a test of equality of regression intercepts across identifiable subgroups (Gulliksen & Wilks, 1950).

Test of Equality of Standard Errors of Estimate

The standard error of estimate measures the amount of error involved in the prediction of the criterion (in this case first-year medical school GPA) from the predictor (Jensen, 1980). Reynolds (1982) defined the standard error of estimate as the standard deviation of the errors (or residuals) observed when predicting the criterion from the predictor within a sample. Reynolds further explained that a relationship exists between the standard error of estimate and the correlations between the predictor and the criterion for two different groups; if the correlations between the predictor and the criterion for the different subgroups are identical, then the standard errors of estimate for the subgroups should be equal. Also the magnitude of the standard error of estimate is based on the spread of the observed criterion scores around the predicted criterion scores. A consistently larger standard error of estimate for one group compared to another indicates a wider spread of the observed scores from the predicted scores and hence a less precise prediction for that group.

According to Reynolds (1982) the standard error of estimate across subgroups can be compared using an F ratio test:

$$F = \frac{SE_1^2}{SE_2^2},\ df = N_1\text{-2 and } N_2\text{-2}$$

SE_1^2 and SE_2^2 represent the square of the standard errors of estimate for group 1 and group 2, respectively.

Test of Equality of Regression Slopes and Equality of Regression Intercepts

When a test score predicts future performance similarly across the different subgroups, it implies that the regression slopes and regression intercepts must be the same across each subgroup. Reynolds (1982) referred to this situation as homogeneity of regression across groups or nonbiased prediction. Alternatively, whenever the slopes or intercepts differ significantly across the subgroups prediction bias occurs and thus a single regression equation based on the combined group should not be used to predict performance for members in the different subgroups (Linn, 1973). Rather, separate equations must be used for each group for fairness in prediction to occur" (Reynolds, 1982, p. 216).

In this study, ANCOVA test of homogeneity or equality of regression coefficients was employed to determine whether the slopes and intercepts of the regression equations for the different subgroups of examinees are significantly different (Zar, 1999). The ANCOVA test was performed by using the Chow test or procedure.

The Chow test was developed by econometricians for the purpose of examining situations such as differences in the earning functions of small firms against large firms, men against women, or among two or more different ethnic groups. The test helps determine whether the coefficients derived from linear regression models are the same for two or more subgroups or for two or more different periods of time (Chow, 1960; Davidson & MacKinnon, 1993). According to Davidson and Mackinnon, such a problem is posed by "partitioning the data into two parts" (p. 375) for the various groups.

> The n-vector y of observations on the dependent variable for the two groups are divided into two vectors, y_1 and y_2, of lengths n_1 and n_2, respectively, and the n * k matrix X of observations on the regressors being divided into two matrices X_1 and X_2, of dimensions n_1 * k and n_2 * k, respectively (Davidson & MacKinnon, 1993, p. 375).

The hypothesis is mathematically written as:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, E\left(uu^{\mathrm{T}}\right) = \delta^2 I,$$

where β_1 and β_2 are each k-vectors of parameter to be estimated. The null hypothesis tested in such cases is that $\beta_1 = \beta_2 = \beta$, reducing the equation to:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \equiv X\beta + u,\ E(uu^{\mathrm{T}}) = \delta^2 I.$$

The unrestricted sum of squared residuals of equation 1 is:

USSR = $y_1^{\mathrm{T}} M_1 y_1 + y_2^{\mathrm{T}} M_2 y_2 = SSR_1 + SSR_2$

where $M_i \equiv I - X_i(X_i^{\mathrm{T}} X_i)^{-1} X_i^{\mathrm{T}}$ for i = 1, 2. Thus the unrestricted sum of squared residuals is the sum of the two sums of squared residuals from the regressions of y_1 on X_1 and y_2 on X_2, respectively.

The restricted sum of squared residuals from equation 2 is:

RSSR = $y^{\mathrm{T}} M_x y$,

where $M_x \equiv I - X(X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}}$

The standard F statistics for the Chow test is then given as:

$$\frac{\left(y^{\mathrm{T}} M_x y - y_1^{\mathrm{T}} M_1 y_1 - y_2^{\mathrm{T}} M_2 y_2\right)/k}{\left(y_1^{\mathrm{T}} M_1 y_1 + y_2^{\mathrm{T}} M_2 y_2\right)/\left(n-2k\right))} = \frac{\left(RSSR - SSR_1 - SSR_2\right)/k}{\left(SSR_1 + SSR_2\right)/\left(n-2k\right)}, \quad d.f. = \text{k and n-2k.}$$

The SPSS procedure for performing the Chow Test is outlined in Appendix J.

Chapter Four: Results

The current study aims at determining the predictive validity of MCAT subtest scores for first-year medical school GPA; undergraduate GPA is included as an additional predictor. Also, the study evaluates differential validity and differential prediction of the MCAT subtest scores for men and women as well as for White, Black, Asian, and Hispanic medical school students.

This chapter presents information regarding the results obtained from relevant statistical analyses performed in order to answer research questions posed in the study. The results are reported in four sections. The first part presents descriptive statistics for the individual MCAT subtest scores, undergraduate science and non-science GPA, and first-year medical school GPA derived both over and within the individual 14 medical schools that agreed to participate in the MCAT Predictive Validity Research. Also, descriptive data are given by examinees' gender and ethnicity. In the second section, predictive validities of the predictors, individually and in combination, are reported. Differential validity of the MCAT subtests across gender and ethnicity are compared in the third section. Finally, differential prediction results derived from Gulliksen and Wilks tests, which involve a comparison of the standard errors of estimate, regression slopes, and regression intercepts based on gender and on ethnicity, are reported. Also, in the final section, patterns of prediction errors or residuals (the difference between the actual and predicted first year medical school GPAs) across the various subgroups are examined.

Descriptive Statistics

Sample sizes for students among the 14 medical schools and within each medical school are presented in Table 4 through Table 7. Table 4 shows the distribution of students for the 1992 and 1993 cohorts in each medical school. In Tables 5, 6, and 7, examinees are further broken down by gender and ethnicity. Of the 3,187 medical students from the 14 medical schools represented across the two years, 1, 646 (51.6 %) students were from the 1992 cohort and the 1993 cohort had 1,541 (48.4 %) students. In each cohort as well as in each school, men and Whites were in the majority. The 1992 cohort consisted of 57.6 % men and 42.4 % women. Ethnicity information for the cohort shows the following percentages: 58.9 % Whites, 11.5 % Blacks, 13.9 % Asians, 6.8 % Hispanic, and 1 % American Indians. The 1993 cohort, on the other hand, comprised of 60.2 % men, 39.8 % women, 62.8 % Whites, 9.7 % Blacks, 12.8% Asians, 5.1 % Hispanic, and 0.9 % American Indians.

Due to the small sample size of American Indians, medical students in this subgroup category were excluded from all ethnicity-related statistical analyses. Similarly, students who failed to report their ethnic status were also excluded from analyses in which knowledge of ethnic group membership was essential in order to answer the research question. Analyses based on ethnicity comparisons were conducted for a school only when at least two ethnic subgroups within that school had sample sizes of at least 30 students per group. A sample size of 30 was considered in the study because based on the Central Limit Theorem as the sample size becomes reasonably large (roughly 30 or more scores) the shape of the sampling distribution of the mean approaches a normal distribution with a population mean of μ and standard deviation of $\sigma/\sqrt{N}$ (Harris, 1998).

Table 4

Number and Percentages of First-Year Medical Students among the 1992 and 1993 Cohorts by School

	1992		1993		1992/1993 Combination	
School	N	%	N	%	N	%
1	102	6.2	107	6.9	209	6.6
2	136	8.3	0	0.0	136	4.3
3	88	5.3	105	6.8	193	6.1
4	241	14.6	279	18.1	520	16.3
5	144	8.7	147	9.5	291	9.1
6	186	11.3	186	12.1	372	11.7
7	123	7.5	139	9.0	262	8.2
8	123	7.5	133	8.6	256	8.0
9	109	6.6	117	7.6	226	7.1
10	92	5.6	96	6.2	188	5.9
11	81	4.9	92	6.0	173	5.4
12	64	3.9	68	4.4	132	4.1
13	68	4.1	72	4.7	140	4.4
14	89	5.4	0	0.0	89	2.8
Total	1,646	51.6	1,541	48.4	3,187	100.0

Table 5

Number and Percentages of Men and Women and White, Black, Asian, and Hispanic First-Year Medical Students among the 1992 and 1993 Cohorts over the 14 Medical Schools

	1992		1993		1992/1993 Combination	
Group	N	%	N	%	N	%
Men	948	57.6	927	60.2	1875	58.8
Women	698	42.4	614	39.8	1312	41.2
White	969	58.9	967	62.8	1936	60.7
Black	189	11.5	149	9.7	338	10.6
Asian	229	13.9	198	12.8	427	13.4
Hispanic	112	6.8	79	5.1	191	6.0
American Indian	17	1.0	14	0.9	31	1.0
Not Indicated	130	7.9	134	8.7	264	8.3
Total	1,646	51.6	1,541	48.4	3,187	100.0

Table 6

Number and Percentages of Men and Women and White, Black, Asian, and Hispanic First-Year Medical Students by School

	Men		Women		Whites		Blacks		Asians		Hispanics	
School	N	%	N	%	N	%	N	%	N	%	N	%
1	147	70.3	62	29.7	157	75.1	30	14.4	13	6.2	4	1.9
2	74	55.4	62	45.6	57	41.9	10	7.4	43	31.6	19	14.0
3	120	62.2	73	37.8	158	81.9	3	1.6	19	9.8	9	4.7
4	325	62.5	195	37.5	268	51.5	87	16.7	96	18.5	54	10.4
5	183	62.9	108	37.1	234	80.4	7	2.4	18	6.2	5	1.7
6	231	62.1	141	37.9	241	64.8	11	3.0	56	15.1	58	15.6
7	155	59.2	107	40.8	146	55.7	46	17.6	40	15.3	23	8.8
8	146	57.0	110	43.0	181	70.7	27	10.5	33	12.9	3	1.2
9	117	51.8	109	48.2	156	69.0	16	7.1	48	21.2	1	.4
10	86	45.7	102	54.3	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
11	93	53.8	80	46.2	108	62.4	14	8.1	32	18.5	8	4.6
12	65	49.2	67	50.8	107	81.1	12	9.1	8	6.1	1	.8
13	92	65.7	48	34.3	117	83.6	7	5.0	10	7.1	3	2.1
14	41	46.1	48	53.9	6	6.7	68	76.4	11	12.4	3	3.4
Total	1,875	58.8	1,312	41.2	1,936	60.7	338	10.6	427	13.4	191	6.0

Note. n.a = Not available.

Table 7

A Break-down of Ethnicity within Gender by School

School	Men				Women			
	White	Black	Asian	Hispanic	White	Black	Asian	Hispanic
1	117	15	8	3	40	15	5	1
2	31	5	25	12	26	5	18	7
3	102	2	8	5	56	1	11	4
4	181	33	72	28	87	54	24	26
5	146	4	12	2	88	3	6	3
6	151	6	36	34	90	5	20	24
7	94	15	28	14	52	31	12	9
8	113	12	10	3	68	15	23	0
9	77	6	30	0	79	10	18	1
10	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
11	59	5	16	5	49	9	16	3
12	56	3	3	0	51	9	5	1
13	79	3	6	2	38	4	4	1
14	4	28	6	2	2	40	5	1
Total	1,210	137	260	110	726	201	167	81

Note. n.a = Not available.

A Chi-square test of association was conducted to determine whether there were significant associations between gender and ethnicity, gender and school, and ethnicity and school for the 1992 and 1993 cohorts. The results revealed significant associations between gender and school for the 1992 and 1993 cohorts, Pearson χ^2 (13, N = 1646) = 27.79, $p < .05$, Cramer's V = .13, and Pearson χ^2 (11, N = 1541) = 31.57, $p < .01$, Cramer's V = .14, respectively. The chi-square test also showed significant association between ethnicity and school for the 1992 cohort, Pearson χ^2 (36, N = 1499) = 561.49, $p < .01$, Cramer's V = .62, and for the 1993 cohort, Pearson χ^2 (30, N = 1393) = 179.81, $p < .01$, Cramer's V = .36, and also between gender and ethnicity for the 1992 cohort, Pearson χ^2 (3, N = 1499) = 24.66, $p < .01$, Cramer's V = .14, and the 1993 cohort, Pearson χ^2 (3, N = 1393) = 34.02, $p < .01$, Cramer's V = .16.

Means and standard deviations for the individual predictors were computed first for the entire study sample by institution (Table 8), and then for the gender and ethnicity subgroups (Table 9). In Table 21 through Table 47 (See Appendix C), means and standard deviations for the predictors and criterion are depicted by school.

The overall mean scores obtained on the individual predictors for the entire sample of medical students were as follows: Biological Sciences, 9.34 (*SD* = 1.91), Physical Sciences, 9.21 (*SD* = 2.14), Verbal Reasoning, 9.33 (*SD* = 1.82), Writing Sample, 6.30 (*SD* = 2.00), undergraduate non-science GPA, 3.53 (*SD* = 0.35), and undergraduate science GPA, 3.36 (*SD* = 0.46).

Table 8

Means and Standard Deviations for MCAT Subtest Scores and Undergraduate GPA by School

	BS		PS		VR		WS		NGPA		SGPA		MGPA	
School	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
1	9.3	1.7	9.4	2.0	9.7	1.7	6.5	2.0	3.4	0.4	3.2	0.5	2.8	0.7
2	10.7	1.9	10.4	2.1	9.5	1.9	6.4	1.9	3.6	0.3	3.6	0.4	79.3	1.1
3	8.7	1.7	8.4	1.7	8.8	1.9	6.2	1.8	3.6	0.4	3.4	0.5	83.6	6.1
4	9.0	2.1	9.0	2.4	9.0	2.0	6.0	2.0	3.5	0.4	3.1	0.5	69.1	9.8
5	8.8	1.5	8.6	1.8	9.3	1.6	6.0	2.1	3.6	0.3	3.4	0.4	3.1	0.6
6	9.2	1.5	9.0	1.7	9.4	1.5	6.2	2.0	3.6	0.3	3.4	0.4	3.1	0.7
7	9.1	1.9	8.8	2.1	8.7	1.8	6.1	1.9	3.5	0.3	3.3	0.4	2.4	0.8
8	10.3	1.8	10.3	2.0	10.0	1.6	6.9	2.1	3.6	0.3	3.5	0.4	91.9	1.9
9	11.4	1.7	11.6	1.8	10.7	1.4	7.1	1.7	3.8	0.2	3.8	0.2	2.6	0.7
10	9.4	1.5	8.9	1.7	9.4	1.5	6.6	2.0	3.6	0.3	3.5	0.4	83.4	5.5
11	9.5	1.6	9.8	1.9	9.5	1.8	6.9	2.0	3.5	0.3	3.4	0.4	2.3	0.8
12	8.8	1.5	8.2	1.4	8. 9	1.6	6.1	1.9	3.6	0.3	3.4	0.4	2.9	0.5
13	8.8	1.5	8.3	1.6	9.1	1.5	5.9	2.1	3.4	0.4	3.3	0.4	2.9	0.7
14	7.1	1.5	7.2	1.4	7.6	2.1	5.2	2.0	3.3	0.4	2.9	0.5	77.8	7.7
Total	9.3	1.9	9.2	2.1	9.3	1.8	6.3	2.0	3.5	0.4	3.4	0.5		
Median	9.0		9.0		9.0		6.0		3.6		3.4			

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA = Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 9

Means and Standard Deviations for MCAT Subtest Scores and Undergraduate GPA by Gender and Ethnicity

	BS		PS		VR		WS		NGPA		SGPA	
Group	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Gender												
Men	9.65	1.84	9.69	2.13	9.34	1.82	6.22	2.01	3.50	.36	3.37	.43
Women	8.91	1.94	8.53	1.97	9.32	1.82	6.42	1.97	3.57	.33	3.34	.49
Ethnicity												
White	9.65	1.68	9.55	1.97	9.77	1.55	6.45	1.98	3.58	.32	3.46	.37
Black	7.07	1.63	6.90	1.50	7.35	1.92	5.45	1.90	3.25	.39	2.80	.52
Asian	10.14	1.75	10.22	1.97	9.34	1.67	6.42	1.95	3.62	.27	3.47	.38
Hispanic	8.20	1.83	7.82	1.91	8.19	1.95	5.81	1.97	3.33	.38	3.01	.50

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA = Undergraduate non-science GPA; SGPA = Undergraduate science GPA; FYGPA= First-year medical school GPA.

A summary of median values for the predictors showed that three of the MCAT subtest scores (Biological Sciences, Physical Sciences, and Verbal Reasoning) had the same median score, 9.0. The median score for the Writing Sample subtest was 6.0 and for undergraduate GPAs, 3.6 and 3.4 for non-science and science GPA, respectively.

School 9 had the highest mean scores across all of the 4 MCAT subtests and for both non-science and undergraduate science GPAs. School 14 had the lowest mean scores across the predictors. Descriptive information based on gender for the pooled data indicated that, in general, men obtained higher group means on the Biological Sciences (*M* = 9.65, *SD* = 1.84) and Physical Sciences (*M* = 9.69, *SD* = 2.13) tests than women, who obtained mean scores of 8.91 (*SD* = 1.94) and 8.53 (*SD* = 1.97) on the Biological Sciences and Physical tests, respectively. Mean scores on the Verbal Reasoning test were slightly higher for men (*M* = 9.34, *SD* = 1.82) compared to women (*M* = 9.32, *SD* = 1.82). Women, however, slightly outscored men on the Writing Sample test with a mean of 6.42 (*SD* = 1.97) versus a mean 6.22 (*SD* = 2.02) for men. With respect to undergraduate non-science and science GPAs, men had slightly higher mean science GPAs than women while women had higher non-science GPAs than men.

Among the ethnic groups, Asians outperformed White, Black, and Hispanic students on the sciences subtests with mean scores of 10.14 (*SD* = 1.76) on the Biological Sciences test and 10.22 (*SD* = 1.97) on the Physical Sciences test. Compared to the other ethnic groups, Whites obtained the highest mean scores on the Verbal Reasoning test (*M* = 9.77, *SD* = 1.55) and on the Writing Sample test (*M* = 6.45, *SD* = 1.98). Hispanics and Blacks obtained lower mean scores across all the MCAT subtests compared to Whites and Asians, with Blacks typically having the lowest mean scores across all the MCAT subtests and for both undergraduate science and non-science GPAs.

First-year medical school GPA differed in range for the schools (Table 9). The majority of schools (schools 1, 5, 6, 7, 9, 11, 12, and 13) reported first-year GPAs on a scale of 0 to 4. GPA for the other schools is reported on a 0 to 100 scale (schools 2, 3, 4, 8, 10, and 14).

In most cases, mean scores shown separately for each of the 14 medical schools followed a pattern of men having higher mean scores on both science subtests and slightly higher mean scores on the Verbal Reasoning. Men also had higher undergraduate science GPAs, and first-year medical school GPAs. Women, on the other hand, performed better on the Writing Sample tests and had higher non-science GPAs.

Across ethnicity, scores on the sciences subtests were higher for White and Asian students. Also, Whites obtain mostly higher mean scores on the Verbal Reasoning and Writing Sample tests as well as on undergraduate and first-year medical school GPA. In general, Black students typically had the lowest mean scores across MCAT subtests, undergraduate GPA, and first-year medical school GPA compared to the students in the other ethnic subgroups.

Predictive Validity

Intercorrelations between the individual predictors and the criterion variable are reported by institution, and then by gender and ethnicity for each school (See Table 48 through Table 88 in Appendix D).

Simple and multiple linear regression analyses were applied to the data to obtain zero-order and multiple correlations between the predictors and the criterion. Prior to performing the regression analyses, the data were tested to find out whether they met the assumptions of regression.

Linearity was assessed through the examination of the intercorrelations between the predictors and the criterion, scatter plots, and standardized residual plots, which showed that there was a straight line relationship between the predictors and the criterion as well as between the standardized predicted scores and standardized residual/error values. The results indicated that the data met the assumption of linearity.

Normality was evaluated through the assessment of histograms and Kolmogorov-Smirnov statistics. The Kolmogorov-Smirnov statistic tests the null hypothesis that the population scores are normally distributed. The null hypothesis was retained, implying that the assumption of normality was met.

Predictive Validities of the Individual Predictors

Validity coefficients for the individual predictors are shown in Table 10. Due to the unavailability of information on the unselected examinees, validity coefficients were uncorrected for range restriction and criterion unreliability. The results revealed variations in the validity coefficients obtained for each predictor across the 14 medical schools.

Table 10

Validity Coefficients for the Individual Predictors

School	N	BS	PS	VR	WS	NGPA	SGPA
1	196	.20	.22	.21	.11	.20	.37
2	111	.63	.61	.57	.17	.47	.52
3	190	.32	.33	.26	.12	.31	.43
4	238	.65	.54	.35	.11	.35	.50
5	283	.25	.27	.05	.05	.17	.30
6	365	.39	.30	.16	.14	.12	.30
7	256	.59	.46	.33	.22	.29	.43
8	253	.44	.40	.23	.13	.37	.52
9	224	.16	.18	.01	.08	.13	.30
10	180	.30	.29	.19	.13	.09	.36
11	163	.39	.42	.21	.11	.25	.36
12	131	.10	.02	.20	.09	.30	.30
13	134	.21	.07	.09	.06	.04	.23
14	88	.32	.36	.06	.09	.04	.18
Range		.10 -.65	.02 - .61	.01 - .57	.05 -.22	.04 -.47	.18 -.52
Median		.32	.32	.21	.11	.23	.36

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA = Undergraduate non-science GPA; SGPA = Undergraduate science GPA.

The median validity coefficient for the Biological Sciences subtest across all 14 medical schools was .32. School 2 obtained the highest validity coefficient, r (111) = .63, $p < .05$, when the Biological Sciences subtest was used as a predictor of first-year medical school GPA. The results for school 2 imply that about 40% of the variance in first-year medical school GPA was explained as a result of using the Biological Sciences subtest alone.

Validity coefficients obtained for the Physical Sciences subtest ranged from .02 to .61 for the 14 medical schools. Schools 12 and 13 had the lowest validity coefficients, .02 and .07, respectively. School 9 had the next lowest validity coefficient, r (224) = .18, $p < .01$, which means that only 3% of the variance in first-year medical school GPA was explained by Physical Sciences subtest scores alone. The median validity coefficient for the Physical Sciences subtests across all 14 medical schools was .32.

The validity coefficients for the Verbal Reasoning subtest were generally lower than those obtained for the science subtests. Implying that the strength of the relationship between first-year medical school GPA and the Verbal Reasoning test was of a lesser magnitude compared to the criterion's relationship with the sciences subtests. The median validity coefficient for the Verbal Reasoning subtest was .21.

The Writing Sample subtest had the lowest validity coefficients across the schools compared to the other MCAT subtests. The median validity coefficient obtained across all of the 14 medical schools was .11.

Generally, undergraduate science GPA showed higher validity coefficients than undergraduate non-science GPA. The median validity coefficients across the medical schools for undergraduate science and non-science GPAs were .36 and .23, respectively.

In summary, the median results indicate that when scores on each of the MCAT subtests, that is, the Biological Sciences, Physical Sciences, Verbal Reasoning, and Writing Sample tests, were considered alone, they accounted for 10.24 %, 10.24 %, 4.41 %, and 1.21 % of the variance in first-year medical school GPA, respectively. On the other hand, undergraduate non-science GPA alone, accounted for approximately 5% of the variance in first-year medical school GPA while undergraduate science GPA alone, accounted for 12.96 % of the variance in first-year medical school GPA.

Predictive Validities of the Combinations of Predictors

In this study, the predictive validities of 6 different combinations of predictors, namely, the combination of the individual MCAT subtest scores, the combination of undergraduate science and non-science GPA, the combination of the MCAT subtest scores and undergraduate GPA, MCAT composite scores, a subset of the Biological Sciences, Physical Sciences, and Verbal Reasoning tests with a subset of the Writing Sample subtest, and the combination of MCAT composite scores and undergraduate GPA, were examined. Tolerance and variance inflation factor values for the predictors were evaluated to help identify the possibility of the occurrence of multicollinearity; that is to make sure that there was no collinearity among the predictors and that none of the predictors were linear functions of the others. Tolerance values for predictors were typically greater than 0.1; a point below which tolerance is considered a cause for concern (Norusis, 1998). Variance inflation factor values were typically less than 10; a rule of thumb which indicates that variance inflation factors are generally a cause for concern (Stevens, 1996). Results of validity coefficients obtained for the combinations of predictors are shown in Table 11.

Table 11

Validity Coefficients for the Combinations of Predictors (Stepwise Regression Models)

School	MCATs	UGPA	MCATs, UGPA	MCAT composite	BS, PS, VR, with WS	MCAT composite, UGPA
1	.22	.37	.39	.27	.22	.39
2	.70	.52	.72	.64	.70	.69
3	.37	.43	.48	.38	.38	.49
4	.66	.50	.68	.59	.62	.64
5	.30	.30	.41	.23	.26	.37
6	.41	.30	.49	.37	.37	.47
7	.59	.43	.63	.53	.55	.59
8	.47	.52	.58	.41	.44	.56
9	.18	.30	.30	.10	.16	.30
10	.35	.37	.48	.34	.35	.48
11	.46	.36	.54	.44	.45	.54
12	.20	.30	.30	.17	.15	.30
13	.21	.23	.31	.18	.18	.28
14	.36	.20	.36	.29	.31	.39
Range	.18 - .70	.20 - .52	.30 - .72	.10 - .64	.15 - .70	.28 - .69
Median	.37	.37	.48	.36	.36	.48

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; MCATs = Combination of BS, PS, VR, and WS; UGPA = Combination of Non-Science and Undergraduate science GPA.

Stepwise regression analyses were performed to determine multiple correlations between first-year medical school GPA and the combinations of MCAT subtest scores, the combination of undergraduate science and non-science GPA, the combination of the MCAT subtest scores and undergraduate GPA, a subset of the Biological Sciences, Physical Sciences, and Verbal Reasoning tests with a subset of the Writing Sample subtest, and the combination of MCAT composite scores and undergraduate GPA. This method of model selection was used to examine whether there was a consistent pattern by which the predictors entered the regression models for the individual medical schools.

The results indicated that, when Biological Sciences, Physical Sciences, Verbal Reasoning, and Writing Sample test scores were allowed to enter into stepwise regression analyses, only scores on the Biological Sciences, Physical Sciences, and occasionally Verbal Reasoning subtests, entered the model at the .05 level of significance for the majority of the medical schools. A similar pattern was observed across the 14 medical schools when the subset of Biological Sciences, Physical Sciences, and Verbal Reasoning subtests with a subset of Writing Sample test predictor set was considered. On the other hand, only undergraduate science GPA was entered at the .05 level of significance when the undergraduate science and non-science GPAs were considered as predictors of first-year medical school GPA.

For the combination of MCAT subtest scores and undergraduate GPAs, a variation of results was observed. Regression results for the five of medical schools (schools 4, 6, 7, 8, and 13) revealed that only Biological Sciences subtest scores and undergraduate science GPA were entered into the regression model ($p < .05$). Also, for schools 5 and 10, only undergraduate science GPA, Biological Sciences, and Physical

Sciences entered the regression equation ($p < .05$). Schools having unique regression models were schools 1, 2, 3, 9, 11, 12, and 14.

In school 1, only undergraduate science GPA and the Verbal Reasoning subtest were entered into the regression model at the .05 level of significance. Step 1 of the regression analysis had a multiple correlation of .37 and R^2 of .13 ($p < .01$), representing the proportion of variance that can be explained by undergraduate science GPA. At step 2, the multiple correlation was .39 and R^2 of .15 ($p < .01$), representing the proportion of variance that can be explained by undergraduate science GPA plus Verbal Reasoning.

Alternatively, Biological Sciences, Verbal Reasoning, and undergraduate science GPA were entered into the regression equation for school 2. The Biological Sciences subtest entered the regression equation first, R (111) = .63 and $R^2 = .39$ ($p < .01$). The Verbal Reasoning subtest was the next to enter the regression model. The Biological Science plus the Verbal Reasoning subtests had a multiple correlation of .68 ($p < .01$), explaining about 46 % of the variance in first-year medical school GPA. At the next step, step 3, undergraduate science GPA entered the regression model. The regression model with the Biological Sciences subtest, Verbal Reasoning subtest, and undergraduate science GPA had a multiple correlation of .72 ($p < .01$).

Only undergraduate science GPA and Physical Sciences were entered into the stepwise regression model at the .05 level of significance in school 3. The first step had a multiple correlation of .43 ($p < .01$) with undergraduate science GPA alone in the regression equation. The Physical Science subtest then entered the model (R (190) = .48, $p < .01$). Whilst in school 9, only undergraduate science GPA entered the model (R (224) = .30, $p < .01$).

For school 11, Physical Science entered the model at step 1, with a multiple correlation of .42 ($p < .01$). In step 2, science undergraduate entered the model (R (163) = .51, $p < .01$), and in step 3, Verbal Reasoning entered the regression model, R (163) = .54 and $R^2 = .29$ ($p < .01$), representing the proportion of variance in first-year medical school GPA that can be explained by Physical Sciences, undergraduate science GPA, and Verbal Reasoning.

School 12 was the only school that had undergraduate non-science GPA enter its stepwise regression model, R (131) = .30 and $R^2 = .09$ ($p < .01$), and in school 14, only Physical Sciences was entered the regression equation, R (88) = .36, $p < .01$.

For the MCAT composite scores, a sum of scores on the individual MCAT subtests was computed and the result was then entered into full regression models. Correlations obtained from this predictor set were evidently lower than those obtained when the individual MCAT subtest scores were considered in combination for almost all the institutions. However, after using Fisher's z transformation to compare validity coefficients obtained for the combination of the individual MCAT subtest scores and for the MCAT composite scores for the individual schools it was observed that the correlations obtained from the 2 predictor sets were not significantly different ($p > .05$).

Multiple regression with blockwise selection was completed to determine the incremental effects of adding the block of the Writing Sample subtest to models already including the block of Biological Sciences, Physical Sciences, and Verbal Reasoning subtest scores. Writing Sample subtest scores, however, failed to enter the regression models for all 14 medical schools.

The ranges of the multiple correlations between the combinations of the individual MCAT subtest scores, undergraduate science and non-science GPA, individual MCAT subtest scores plus undergraduate GPA, and the MCAT composite plus undergraduate GPA with first-year medical school GPA were .18 to .70, .20 to .52, .30 to .72, and .28 to .69, respectively. Median multiple correlations across all 14 medical schools ranged from .37 to .48 for the combinations of predictors. These results indicate that approximately 14 % and 23% of the variance in first-year medical school GPA can be accounted for by the combination of MCAT subtest scores and the combination of MCAT subtest scores with undergraduate GPA, respectively.

Due to the inconsistency of the patterns of entry of predictors into the stepwise regression models, full models were also developed whereby all predictors were entered simultaneously into the prediction models. Validity coefficients obtained with the full model approach are depicted in Table 12.

The ranges of the multiple correlations between the combinations of the individual MCAT subtest scores, undergraduate science and non-science GPA, individual MCAT subtest scores plus undergraduate GPA, and the MCAT composite plus undergraduate GPA with first-year medical school GPA using the full model were .18 to .70, .20 to .52, .30 to .72, and .29 to .69, respectively. Median multiple correlations across all 14 medical schools ranged from .39 to .49 for the combinations of predictors.

Table 12

Validity Coefficients for the Combinations of Predictors (Full Regression Models)

School	MCATs	UGPA	MCATs, UGPA	MCAT composite	BS, PS, VR, with WS	MCAT composite, UGPA
1	.27	.37	.41	.27	.22	.40
2	.71	.54	.73	.64	.70	.69
3	.39	.43	.50	.38	.39	.49
4	.66	.51	.69	.59	.62	.64
5	.30	.30	.41	.23	.26	.37
6	.41	.31	.50	.37	.37	.47
7	.60	.44	.64	.53	.55	.60
8	.47	.52	.59	.41	.44	.56
9	.23	.30	.33	.10	.16	.30
10	.37	.37	.49	.34	.35	.48
11	.48	.36	.56	.44	.45	.54
12	.23	.33	.38	.17	.15	.35
13	.22	.25	.34	.18	.18	.29
14	.42	.20	.46	.29	.31	.40
Range	.22 - .71	.20 - .54	.33 - .73	.10 - .64	.15 - .70	.29-.69
Median	.39	.37	.49	.36	.36	.48

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; MCATs = Combination of BS, PS, VR, and WS; UGPA = Combination of Non-Science and Undergraduate science GPA.

These results indicate that approximately 15 % and 24% of the variance in first-year medical school GPA can be accounted for by the combination of MCAT subtest scores and the MCAT subtest scores with undergraduate GPA, respectively. Again, using Fisher's *z* transformation to compare validity coefficients obtained for the combination of the individual MCAT subtest scores against validity coefficients obtained for the MCAT composite scores for the individual schools produced significant results for only 1 medical school, school 9, z (224) = 1.99, $p < .05$.

Differential Validity

In this section, validity coefficients obtained for the individual MCAT subtests are examined for evidence of differential validity between men and women and among White, Black, Asian, and Hispanic medical students.

Comparison of Validity Coefficients across Gender

Table 13 shows validity data for the individual MCAT subtest scores and for the combination of the MCAT subtest scores by gender. In general, validity coefficients appeared to be higher for women than for men. However, using Fisher's z transformations to compare validity coefficients for the different gender subgroups revealed that validity coefficients obtained for the Biological Sciences subtest were significantly different for men and women for only 2 schools, school 5, $z = -2.15$, $p < .05$, and school 13, $z = -2.54$, $p < .05$. The validity coefficients were, in both cases, higher for women than for men. In school 5, the validity coefficient for women was .38 compared to .13 for men, and in school 13, .49 compared to .05 for men.

Table 13

Validity Coefficients for Men versus Women by School

	Men					Women				
School	BS	PS	VR	WS	MCATs	BS	PS	VR	WS	MCATs
1	.20	.25	.24	.18	.30	.19	.15	.15	-.05	.58
2	.55	.58	.52	.14	.58	.69	.63	.65	.18	.74
3	.27	.25	.26	.11	.27	.39	.48	.26	.18	.48
4	.62	.51	.33	.22	.62	.66	.57	.36	-.07	.70
5	.13	.20	.08	.06	.20	.38	.30	.08	.04	.38
6	.32	.27	.16	.11	.35	.47	.28	.18	.21	.47
7	.48	.28	.28	.18	.48	.64	.60	.34	.24	.69
8	.44	.40	.22	.13	.47	.43	.41	.23	.14	.43
9	.15	.12	-.04	-.06	n.c.	.18	.25	.03	-.10	.25
10	.41	.35	.18	.10	.41	.22	.29	.21	.17	.29
11	.29	.37	.15	.17	.37	.47	.54	.28	.06	.58
12	.06	-.01	.16	-.01	n.c.	.18	-.03	.34	.40	.47
13	.05	-.11	.04	-.17	n.c.	.49	.48	.32	.16	.60
14	.55	.47	.31	.11	.55	.17	.29	.20	.07	.42

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; MCATs = Combination of BS, PS, VR, and WS; UGPA = Combination of Non-Science and Undergraduate science GPA. n.c. = Not computed; none of the predictors were entered into the regression model.

For the Physical Sciences, only school 7, $z = -3.194$, $p < .05$, and school 13, $z = -2.17$, $p < .05$, had significantly different validity coefficients for men and women; again, with women showing more validity compared to men.

Gender differences for the other subtests, Verbal Reasoning and Writing Sample, failed to reveal significant results across all of the 14 medical schools. Also, when the subtests were considered in combination, none of the validities were significantly different for men and women.

Comparison of Validity Coefficients across Ethnicity

Ethnicity comparisons were performed for 8 schools (schools 1, 2, 4, 6, 7, 8, 9, and 11); 3 comparisons between Whites and Blacks, 7 comparisons between Whites and Asians, and 2 comparisons between Whites and Hispanics. Table 14 shows the validity coefficients for White versus Black medical students. Similarly, Tables 15 and 16 show the validity coefficients obtained for Whites versus Asians and Whites versus Hispanics, respectively.

Comparisons for the validity coefficients derived for the Biological Sciences, Physical Sciences, Verbal Reasoning, and Writing Sample subtest scores indicated no significant differences among the different ethnic subgroups at the .05 level of significance. However, a significant difference occurred between Whites and Blacks in school 1, $z = -2.29$, $p < .05$ for the combination of subtest scores.

Table 14

Validity Coefficients for Whites versus Blacks by School

	Whites					Blacks				
School	BS	PS	VR	WS	MCATs	BS	PS	VR	WS	MCATs
1	.21	.29	.24	.16	.33	.17	.02	.04	-.33	.70
4	.53	.38	.03	.08	.53	.40	.25	.02	.20	.50
7	.35	.19	.12	.15	.35	.57	.27	.03	-.06	.57

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; MCATs = Combination of BS, PS, VR, and WS; UGPA = Combination of Non-Science and Undergraduate science GPA.

Table 15

Validity Coefficients for Whites versus Asians by School

	Whites					Asians				
School	BS	PS	VR	WS	MCATs	BS	PS	VR	WS	MCATs
2	.39	.33	.15	.18	.39	.42	.33	.13	-.52	.64
4	.53	.38	.03	.08	.53	.59	.21	.05	.02	.59
6	.32	.22	-.002	.13	.32	.13	-.04	.12	.05	n.c.
7	.35	.19	.12	.15	.35	.50	.19	.08	-.05	.50
8	.33	.29	-.02	.14	.33	.42	.25	-.23	-.24	.42
9	.02	.11	.11	-.08	n.c.	.25	.25	.17	-.19	n.c.
11	.16	.23	.18	.11	.26	.48	.46	-.04	-.06	.58

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; MCATs = Combination of BS, PS, VR, and WS; UGPA = Combination of Non-Science and Undergraduate science GPA.

n.c. = Not computed; none of the predictors were entered into the regression model.

Table 16

Validity Coefficients for Whites versus Hispanics by School

	Whites					Hispanics				
School	BS	PS	VR	WS	MCATs	BS	PS	VR	WS	MCATs
4	.53	.38	.03	.08	.53	.39	.57	.33	.03	.57
6	.32	.22	-.002	.13	.32	.33	.39	.24	.004	.39

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; MCATs = Combination of BS, PS, VR, and WS; UGPA = Combination of Non-Science and Undergraduate science GPA.

Differential Prediction

Gulliksen and Wilks Tests for Gender and Ethnicity

Standard errors of estimate, regression coefficients, and regression intercepts for men and women, and for White, Black, Asian, and Hispanic students in each of the 14 medical schools are presented in Tables 89 through 124 (Appendix E).

Gulliksen and Wilks Tests by Gender.

Results from Gulliksen and Wilks tests for men and women are shown in Table 17. A total of 70 men/women comparisons were made for the 14 medical schools.

Standard errors of estimate were found to be significantly different ($p < .05$) for about one-fifth (13) of the 70 men/women comparisons. These were for the Biological Sciences test, schools 1, 5, 6, 7, 9, 11, 12, and 13, and for the Physical Sciences subtest and the combination of MCAT subtest scores, school 7 and school 12, and lastly for the Verbal Reasoning and Writing Sample subtests, school 12.

Out of the 70 comparisons of regression slopes and intercepts, 7 regression slopes and 7 regression intercepts showed significant differences across gender. However, the size of the effect were quite small for all the slope and intercept differences.

Table 17

Results from Gulliksen and Wilks Regression Tests by Gender

	BS			PS			VR			WS			MCATs		
School	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0
1	*	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	nc	nc	nc
2	ns	*	*	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
4	ns	ns	ns	ns	*	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
5	*	ns	*	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
6	*	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
7	*	ns	ns	*	**	**	ns	ns	ns	ns	ns	ns	*	ns	ns
8	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
9	*	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	nc	nc	nc
10	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
11	*	ns	ns	ns	*	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
12	*	ns	ns	*	ns	ns	*	ns	ns	*	ns	*	nc	nc	nc
13	*	*	ns	ns	**	**	ns	ns	ns	ns	ns	ns	nc	nc	nc
14	ns	ns	*	ns	ns	ns	ns	*	*	ns	ns	ns	ns	ns	ns

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept

ns = Not significant; nc = Not computed; none of the predictors were entered into the regression model.

*$p < .05$. ** $p < .01$.

Gulliksen and Wilks Tests by Ethnicity.

Results from Gulliksen and Wilks tests for Whites, Blacks, Asians, and Hispanics are depicted in Table 18 through Table 20. A total of 35 White/Asian comparisons were made for 7 medical schools. Also, 15 White/Black comparisons were made for 3 medical schools and 10 White/Hispanic comparisons were made for 2 medical schools.

Ethnicity-based comparisons for standard errors of estimate showed that 2 of the 35 White/Asian comparisons were significantly different ($p < .05$). These were for the Biological Sciences test for schools 11, and again, for the combination of MCAT subtest scores for school 11.

Further comparisons of regression slopes and intercepts revealed relatively fewer slope differences and a moderately higher intercept differences than that obtained for gender. Only 2 (1 White/Black and 1 White/Asian) regression slopes showed significant differences. On the other hand, 11 (4 White/Black, 6 White/Hispanic, and 1 White/Asian) intercepts were significantly different. The sizes of the effect for these differences were, however, quite small in magnitude.

Table 18

Results from Gulliksen and Wilks Regression Tests by Ethnicity (White/Black)

	BS			PS			VR			WS			MCATs		
School	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0
1	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	*	*	nc	nc	nc
4	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	**	ns	ns	ns
7	ns	ns	*	ns	ns	*	ns	ns	ns	ns	ns	ns	ns	ns	ns

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept

ns = Not significant; nc = Not computed; none of the predictors were entered into the regression model.

*p < .05. ** p < .01.

Table 19

Results from Gulliksen and Wilks Regression Tests by Ethnicity (White/Asian)

	BS			PS			VR			WS			MCATs		
School	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0
2	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	**	**	ns	ns	ns
4	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
6	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	nc	nc	nc
7	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
8	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
9	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	nc	nc	nc
11	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept

ns = Not significant; nc = Not computed; none of the predictors were entered into the regression model.

** $p < .01$.

Table 20

Results from Gulliksen and Wilks Regression Tests by Ethnicity (White/Hispanic)

	BS			PS			VR			WS			MCATs		
School	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0
4	ns	ns	ns	ns	ns	*	ns	ns	*	ns	ns	ns	ns	ns	*
6	ns	ns	ns	ns	ns	*	ns	ns	*	ns	ns	ns	ns	ns	*

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept

ns = Not significant.

*$p < .05$.

Residual Analysis

First-year medical school GPA was regressed on MCAT subtest scores, individually and in combination, for the total sample of students within each school. Predicted criterion scores were then calculated for each subtest and for the combination of subtests for the various gender and ethnicity subgroups. Mean residuals were derived by subtracting predicted scores from actual scores obtained on the criterion. A positive value indicated underprediction of performance in the first year of medical school; likewise a negative value represented overprediciton of performance in the first year of medical school.

The extent of over- or underprediction was examined by gender and ethnicity within each school and over the 14 medical schools. Differences between the actual and predicted first-year medical school GPAs for men and women and for Whites, Blacks, Asians, and Hispanics are depicted in Tables 125 to 128 (See Appendix F). Also, the number and percentage of students whose first-year medical school GPAs were over- or underpredicted by MCAT subtest scores, individually and in combination, are also presented in Table 129 through Table 148 (Appendix G).

Residual Analysis by Gender.

A summary of the data showed that the direction of the mean residuals across the MCAT subtests for men and women were not consistent within the medical schools. Weighted mean residuals for men over the 14 medical schools, however, were .03 for Biological Sciences, -.08 for Physical Sciences, .17 for Verbal Reasoning, .18 for Writing Sample, and -.09 for the combination of MCAT subtests. The data indicated that on the average, men’s first-year medical school GPAs were underpredicted by Biological

Sciences, Verbal Reasoning, and the Writing Sample subtests alone, and overpredicted by the Physical Sciences subtest alone and by the combination of MCAT subtests. The underprediction for men was largest for the Writing Sample subtest alone. On the other hand, the overprediction for men was largest for the combination of MCAT subtests.

Examination of the weighted mean residuals derived for women (.09 for Biological Sciences, .11 for Physical Sciences, -.22 for Verbal Reasoning, -.28 for the Writing Sample subtest, and .13 for the combination of MCAT subtest scores) over the 14 medical schools showed that women's first-year medical school GPAs were underpredicted by the Biological Sciences and Physical Sciences subtests alone, and the combination of MCAT subtests; and overpredicted by the Verbal Reasoning and Writing Sample subtests alone. The underprediction for women was largest for the combination of MCAT subtests, while overprediction was largest for the Writing Sample subtest alone.

Further examination of the number and percentages of men whose first-year medical school GPAs were either over- or underpredicted by using common regression equations derived for men and women, again, revealed inconsistencies within the schools. The number and percentages over the 14 medical schools showed that the criterion scores for the majority of men or for larger percentages of men were underpredicted than overpredicted by the Biological Sciences, Verbal Reasoning, and Writing Sample subtests; likewise criterion scores for larger percentages of men were overpredicted than underpredicted by the Physical Sciences subtest and the combination of MCAT subtest scores. Alternatively, the number and percentage of women whose first-year medical school GPAs were either over- or underpredicted by using common regression equations derived for men and women showed that the criterion scores for the majority of women were underpredicted than overpredicted by the Biological Sciences, Physical Sciences, and combination of MCAT subtest scores; and overpredicted than underpredicted by the Verbal Reasoning and Writing Sample subtests over the 14 medical schools.

Residual Analysis by Ethnicity.

Weighted mean differences between actual and predicted first-year medical school performance for White, Black, and Hispanic students showed a consistently positive pattern for Whites across all of the MCAT subtests, individually and collectively, and a consistently negative pattern for Black and Hispanic students across all of the MCAT subtests, individually and collectively. The results imply that first-year grades were typically underpredicted for Whites and overpredicted for Blacks and Hispanics by the MCAT subtest scores, individually and collectively. In other words, White medical students tended to performed better in the first-year of medical school than predicted by the individual MCAT subtests alone and by the combination of MCAT subtest scores, while Blacks and Hispanics did not perform as well as predicted by the MCAT subtest scores, individually and in combination. The degrees of overprediction over the schools were constantly larger for Black students while Hispanics had a consistently higher percentage of students whose grades were overpredicted by the MCAT subtest scores.

The weighted residuals for Asians (-.04 for Biological Sciences, -.03 for Physical Sciences, .04 for Verbal Reasoning, .56 for the Writing Sample subtest, and -.20 for the combination of MCAT subtest scores), however, failed to show a consistent pattern for the MCAT subtest scores. The data showed that on the average first-year medical school GPAs for Asian medical students were underpredicted by the Verbal Reasoning and

Writing Sample subtests alone. In addition, first-year medical school GPAs were overpredicted by Biological Sciences and Physical Sciences subtests alone, and by the combination of MCAT subtest scores. The degree of underprediction was largest for the Writing Sample subtest alone and the degree of overprediction was largest for the combination of MCAT subtests.

Chapter Five: Summary, Interpretation, Conclusions, and Recommendations

This chapter provides a summary of the dissertation and interpretation of the research findings. Also, conclusions derived from the findings are reported. Finally, the limitations of the study are detailed and recommendations are made for further research.

Summary

The admissions process is one of crucial importance to any academic institution. Admissions committee members are especially interested in knowing that they made the most valid decisions during the course of the admissions process; that is, that the students they selected for admissions do indeed fit well with their programs of study. An important question that is often asked is whether matriculants to a program are likely to do well in a department's academic curriculum (Willingham, Lewis, Morgan, & Ramist, 1990).

In order to ensure the validity of their decisions, admissions committees outline a number of items they consider as being important to the admissions process. Although the criteria for admissions vary among colleges, two selection tools, students' academic records and scores on standardized tests, have been identified as criteria which carry the most weight in the selection process. Academic records provide admission committees with a sense of students' performance or achievement levels on previous programs while test scores, on the other hand, provide standardized information regarding applicants' academic abilities.

"Any measure that is commonly used for selection has effects throughout the educational system, both direct and indirect. There is, therefore, a need to validate or justify that use." (Willingham, Lewis, Morgan, & Ramist, 1990, p. 6). One area under which such issues fall is predictive validity. Predictive validity studies help to establish information about the usefulness of admissions criteria and often involve an examination of the relationships that exist between selection measures and students' subsequent performance in college.

The purpose of this study was to examine the predictive validity of MCAT subtest scores and undergraduate GPA; two primary selection criteria used in medical school admissions. Specific measures investigated consisted of first-year medical school GPA as the criterion and 12 predictor sets:

1. Biological Sciences scores alone.
2. Verbal Reasoning scores alone.
3. Physical Sciences scores alone.
4. Writing Sample scores alone.
5. Undergraduate non-science GPA alone.
6. Undergraduate science GPA alone.
7. The combination of Biological Sciences, Verbal Reasoning, Physical Sciences, and Writing Sample subtest scores.
8. The combination of undergraduate non-science and science GPA.
9. The combination of Biological Sciences, Verbal Reasoning, Physical Sciences, Writing Sample subtest scores, and undergraduate non-science and science GPA.
10. MCAT total/composite scores.
11. The combination of MCAT total/composite scores, and undergraduate non-science and science GPA.

12. A subset/block of Biological Sciences, Verbal Reasoning, and Physical Sciences subtest scores with a subset/block of Writing Sample subtest scores.

Another aim of the study was to determine whether differential validity and differential prediction of MCAT subtest scores existed across the different gender and ethnic subgroups. With respect to gender and ethnicity, men and Whites constituted the majority groups while the minority groups included women and Blacks, Asians, and Hispanics.

Specific research questions addressed were:

1. How well do undergraduate GPA and MCAT subtest scores, both individually and collectively, predict first-year medical school GPA?
2. How well do MCAT total/composite scores predict first-year medical school GPA?
3. How well do subsets of MCAT subtest scores predict first-year medical school GPA?
4. Are the predictive strengths of MCAT subtest scores, both individually and collectively, consistent across gender?
5. Are the predictive strengths of MCAT subtest scores, both individually and collectively, consistent across ethnicity?

A number of statistical techniques were applied to help answer the research questions; namely, simple and multiple linear regression, Fisher's z transformations, an F-ratio test of equality of standard errors of estimate, and ANCOVA tests of equality of regression slopes and intercepts.

Data from 14 medical schools, which were geographically and ethnically representative of all the medical schools in the U.S., were analyzed. Analyses were conducted separately for each medical institution due to the fact that first-year GPA is school-specific/dependent. Analyses for each school were first performed using the total sample of students within each school. Subsequently, data were analyzed based on gender and ethnicity by school.

Interpretation of the Research Findings

Initially, descriptive data on MCAT subtests scores, undergraduate GPA (both non-science and science GPA), and first-year medical school GPA were provided. The direction of the average MCAT scores by gender indicated that men performed better on the science subtests while women performed better on the writing test. Also, on the average, Asians earned higher scores the science subtests than Whites, Blacks, and Hispanics while Whites performed better on the verbal and writing tests than did minorities. These findings are consistent with results demonstrated in previous studies related to differences in gender and ethnicity performance on the MCAT (Halpern, Haviland, & Killian, 1998; Jones & Mitchell, 1986). The results also show that Asians are the only minority group that, on the average, performs better than the majority group on the subtests of the MCAT.

With regards to undergraduate GPA, women obtained slightly higher mean scores on undergraduate non-science GPA compared to men. However, contrary to conclusions drawn by Willingham and Cole (1997) that women tend to make better grades in school than men, the study results revealed that generally men rather than women had higher mean scores on undergraduate science GPA and on first-year medical school GPA. This may have been due to the fact that men generally perform better on science subjects than women (Azen, Bronner, & Gafni, 2002).

Validity results derived for the individual schools ranged from .10 to .65 for the Biological Sciences test, .02 to .61 for the Physical Sciences subtest, .01 to .57 for the Verbal Reasoning test, and finally .05 to .22 for the Writing Sample subtest. The results showed that the science tests were better predictors of first-year medical school grades (Cullen, Peckham, & Schwarz, 1980; Glaser, Hojat, Veloski, Blacklow, & Goepp, 1992; Ramos, Croen, & Haddow, 1986), followed by the Verbal Reasoning test. Compared to the other three subtests (Biological Sciences, Physical Sciences, and Verbal Reasoning) the Writing Sample subtest had the weakest relationship with first-year medical school GPA. The results may have been due to the fact that in the first year of medical school students have to take basic science courses and their performance in those courses may have no relationship with their communication or writing ability. Second, unlike the other subtests of the MCAT which are multiple-choice tests, the Writing Sample test consists of an essay format thus have may have a lesser reliability.

Ranges in multiple correlations derived from stepwise regressions for the combinations of predictors over the 14 medical schools were .18 to .70 for the combination of MCAT subtest scores, .20 to .52 for the combination of undergraduate non-science and science GPAs, .30 to .72 for the combination of MCAT subtest scores and undergraduate GPAs, and .28 to .69 for the combination of MCAT composite scores and undergraduate GPAs. Multiple correlations ranges derived from full regression models over the medical schools, .22 to .71 for the combination of MCAT subtest scores, .20 to .54 for the combination of undergraduate non-science and science GPAs, .33 to .72

for the combination of MCAT subtest scores and undergraduate GPAs, and .29 to .69 for the combination of MCAT composite scores and undergraduate GPAs, were evidently slightly higher than those obtained from the stepwise regression procedures. However, as expected the multiple correlations obtained from both models were higher than the zero-order correlations between the individual predictors and the criterion.

Validity coefficients for MCAT composite scores were derived from full regression models. Ranges of the correlations obtained from this predictor set were .10 to .64. Although, these correlations appeared to be lower than those obtained for the combination of subtest scores across the 14 medical schools, z transformations later revealed that the correlations were significantly different for only one medical school.

Blockwise selection models were performed for the subset of Biological Sciences, Physical Sciences, and Verbal Reasoning with the subset of Writing Sample subtest scores. The results showed that adding Writing Sample subtest scores to the model that already consisted of Biological Sciences, Physical Sciences, and Verbal Reasoning subtest scores did not significantly increase predictive values for all 14 medical schools.

Predictive validity results for the individual MCAT subtests and undergraduate GPAs, as well as their combinations showed considerable variation across the medical schools that participated in the study. These variations may have been due to differences in the grading criteria among schools (Julian & Lockwood, 2000) and/or differences in the range of talent within each institution (Willingham, Lewis, Morgan, & Ramist, 1990).

On the whole, there appeared to be moderately high correlations between first-year medical school grades and undergraduate GPA and MCAT subtest scores. The results indicated that MCAT subtest scores and previous grades were both, individually and collectively, good predictors of admissions decisions and of medical school freshman grades. Also, as was true in past validity studies (Koenig, Huff, & Julian, 2002; Julian & Lockwood, 2000; Veloski, Callahan, Xu, Hojat, & Nash, 2000; Wiley & Koenig, 1996) the predictive validity data for this study confirmed that the combinations of predictors produce higher validity coefficients than the individual predictors alone.

Differential validity results showed that in the majority of the schools women had higher correlations between MCAT subtest scores, separately and together, and the criterion, than men. After testing for significant gender and ethnicity differences the results revealed differences in validity coefficients between men and women on the Biological and Physical Sciences tests for 2 of the 14 medical schools and between Whites and Blacks on the combination of MCAT subtest scores for 1 medical school, implying the existence of differential validity of MCAT subtest scores across gender and ethnicity. The magnitudes of the correlations between the two sciences subtests and the criterion were higher for women than for men in both schools. The results indicate that there is a stronger relationship between the science subtest scores and first-year medical school GPA for women than for men, which is consistent with the findings by Jones and Vanyur (1985). Likewise the magnitude of the correlation between the combination of MCAT subtest scores and first-year medical school GPA was higher for Blacks than Whites in one school, implying that there was a stronger relationship between the combination of MCAT subtests scores for Blacks than for Whites in that school.

In terms of differential prediction, gender comparisons yielded 13 standard error of estimate, 7 regression slope, and 7 regression intercept differences. On the other hand, ethnicity comparisons revealed 2 standard error of estimate, 2 regression slope, and 11 regression intercept differences. Implying that using common regression equations derived from a pooled sample of men and women and White, Black, Asian, and Hispanic medical students to predict performance in the first year of medical school may result in predictive bias against certain subgroups.

Results from residual analyses later revealed that the grades for men were underpredicted by Biological Sciences, Verbal Reasoning, and the Writing Sample subtests alone, and overpredicted by the Physical Sciences subtest alone and by the combination of MCAT subtests. Alternatively, grades for women were underpredicted by the Biological Sciences and Physical Sciences subtests alone, and the combination of MCAT subtests and overpredicted by the Verbal Reasoning and Writing Sample subtests alone. The results were contrary to that obtained for past gender-related MCAT research which failed to identify the presence of significant gender-related bias in the MCAT (Jones & Vanyur, 1985).

With respect to ethnicity, the study found out that first-year medical school grades for Asians were underpredicted by the Verbal Reasoning and Writing Sample subtests alone. In addition, first-year medical school GPAs were overpredicted by Biological Sciences and Physical Sciences subtests alone, and by the combination of MCAT subtest scores. The results also demonstrated constant underprediction for Whites and consistent overprediction for Blacks and Hispanics across the individual MCAT subtest scores as well as for the combination of MCAT subtest scores. These findings contradict conclusions drawn by Vancover, Reinhart, Solomon, and Haf (1990) that the MCAT are indeed equally predictive for minority and majority groups but are similar to those obtained from an earlier MCAT ethnicity-related study (Koenig, Sireci, & Wiley, 1998) and are in line with the results reported in past ethnicity-related research on other large-scale standardized tests (Breland, 1979; Cleary, 1968; Ramist, 1984; Young, 1991, 1994).

Limitations of the Study

An important limitation of the study is the unavailability of data for each institution's entire applicant pool (that is, data for both the selected and unselected examinees). This information was potentially relevant for correcting derived validity coefficients for restriction of range. Again, the lack of information on important pieces such as the exact medical school course grades or scores obtained by the participants of this study resulted in the researcher's inability to adjust validity coefficients for the effects of criterion unreliability. Thus, although validity coefficients obtained in this study show moderately high relationships between the predictors, MCAT subtest scores and undergraduate GPA, and the criterion, first-year medical school GPA, there is still a possibility that these validity coefficients are underestimations of the true nature of the relationships between the predictors and the criterion. That is, the predictors may be explaining less of the variance in first-year medical school GPA than they actually account for.

Second, due to the limited sample sizes for the minority groups in the individual medical schools, differential validity and differential prediction comparisons based on ethnicity were not compared across all of the 14 medical schools. Ethnic group comparisons were mostly between Whites and Asians. The availability of larger sample sizes within the ethnic subgroups and across medical schools may have illustrated better differential validity and over- and underprediction results.

In addition, minority groups were lumped up into 4 major categories, Whites, Blacks, Asians, and Hispanics, thus eliminating important differences that may have existed between members of the different subcategories. For example Chicanos, Latinos, Puerto Ricans, Cubans, and Mexican Americans were all grouped together as Hispanics, while Chinese, Japanese, Indians, Filipinos, and Vietnamese were all considered Asians. Also, one group, American Indians, was totally eliminated from all ethnicity-related analyses due to their relatively small sample size. This act contributed to the continued lack of validity information for this important subgroup.

Lastly, the data used in the study were obtained from only 14 medical schools and thus the research findings may lack generalizability over all medical schools in the U.S.

Conclusions

In spite of the outlined limitations the researcher is of the opinion that this study is significant because it constitutes one of the few MCAT subtest-specific studies to date. The study contributes to the body of knowledge concerning the predictive validity of MCAT subtest scores and undergraduate GPA and provides a baseline of information for future research. In addition, the research findings presented in the study provide empirical evidence to support the use of MCAT subtests in medical school admissions.

The results of the study establish the fact that MCAT subtest scores and undergraduate GPA, individually and in combination, are valid predictors which have considerable usefulness in predicting first-year medical school grades; the combination of MCAT subtests and undergraduate GPA serve as a more powerful predictor than when either predictor is considered alone. Another important conclusion that can be drawn from the study is that differential validity as well as over- and underprediction of first-year grades occur across examinees' gender and ethnicity. An important point worth noting here, however, is that the patterns of differential validity were very few and the sizes of the effect were mostly small for the medical schools. Also, the patterns of overprediction, and underprediction for men, women, and Asians were inconsistent across all the medical schools, implying that the study failed to identify widespread bias in favor of men and against women and Asians. Constant patterns of underprediction and overprediction were, however, observed across the schools for White, Black, and Hispanic medical students.

Recommendations

Recommendations for Admission Committees

Admission committees may want to consider using the combination of MCAT subtest scores in place of the MCAT composite scores in their prediction equations because it appears that validity coefficients derived from the combination of subtest scores might explain a larger proportion of the variance in first-year medical school GPA than do composite scores.

Second, in this study, stepwise regression was applied with the aim of determining whether a discernible pattern of entry existed for the individual predictors across medical schools. Contrary to expectation, a specific/consistent pattern was not observed. This may have been due to the effect of differences in school-specific factors such as grading. Given that the researcher lacks the theoretical basis to make the argument for the use of stepwise regression, it is recommended that admissions committees consider entering the individual subtests simultaneously into their prediction model, especially since the full model results in slightly higher validity coefficients than those obtained by way of the stepwise regression. Also, admission committee members may want to determine, internally, which of the subtests best predict first-year medical school performance for their students and use these subtests in their prediction models. For example, based on the weak relationship observed between the Writing Sample test and first-year medical school performance, admission committees may decide whether or not they would want to include this subtest in predicting students' performance in medical school.

Third, since differential validity and differential prediction results did not show definitive patterns of gender and ethnicity differences across all of the medical schools included in this investigation, admission committees in the individual institutions may want to conduct their own internal analyses to confirm whether prediction of students' performance differ significantly for the different gender and ethnicity subgroups before they decide to use either total-group or within-group prediction models in their admissions process.

Recommendations for Future Research

Without doubt, this study does not provide all the answers to issues related to the validity of MCAT scores. Clearly, additional research is needed to better understand the nature of the validity of the MCAT.

The usefulness of MCAT subtest scores in predicting other criterion measures of students' performance in medical school, for example, second-year medical school GPA and cumulative first-year and second-year GPA should be explored. In addition, the validity of other predictors, such as students' socio-economic status as well as interview ratings and other non-cognitive factors, which are also considered to be relevant to the medical school admissions process need to be examined in order to present a complete picture of the validity of the decisions made by medical school admissions committees.

Also, an investigation of the causes of between- and within-group differences on the MCAT is another important area that should be pursued. Finally, future research should be conducted on why there is differential validity and differential prediction, and to determine the existence of differential validity and differential prediction for the interaction between gender and ethnicity, especially to determine whether the patterns of underprediction for whites and overprediction for Blacks and Hispanics remain consistent for men and women within these ethnic subgroups.

References

Aiken, L. R. (1994). *Psychological testing and assessment* (8th ed.). Needham Heights, MA: Allyn and Bacon.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillian.

Anderhalter, O. F. (1962). Developing uniform departmental grading standards in a university. *Journal of Experimental Education, 31*, 210-211.

Anderson, M. B., & Swanson, A. G. (1993). Educating medical students: Assessing changes in medical education - the road to implementation. *Academic Medicine, 68* (Suppl.), S1-S46.

Angoff, W. H. (1971). *The College Board Admissions Testing program: Research and development activities relating to the Scholastic Aptitude Test and Achievement Tests*. New York: College Entrance Examination Board.

Angoff, W. H. (1988). Validity: An evolving concept. In H. Wainer & H.I. Braun (Eds.), *Test validity* (pp. 19-30). Hillsdale, NJ: Erlbaum.

Association of American Medical Colleges (1986). *MCAT summary scores report*.

Washington, DC: Author.

Association of American Medical Colleges (2003). *MCAT changes for 2003*. Retrieved March 11, 2004, from http://www.aamc.org/students/applying/advisors/mcat2003changes.pdf

Association of American Medical Colleges (2003). *Curriculum directory*. Retrieved April 7, 2004, from http://services.aamc.org/currdir/section2/start.cfm.

Association of American Medical Colleges (2003). Scoring the MCAT Writing Sample: Examples of MCAT Writing Sample responses and explanations of their scores. Retrieved July 7, 2004, from http://www.aamc.org/students/mcat/studentmanual/writingsample/scoring.pdf.

Austin, D., Clark, V., & Fitchett, G. (1971). *Reading rights for boys: Sex roles in language experiences*. New York: Meredith.

Azen**,** R., Bronner, S., & Gafni, N. (2002). Examination of gender bias in university admissions. *Applied Measurement in Education,* 15, 75-94.

Barzansky, B., & Etzel, S. I. (2003). Educational programs in U.S. medical schools, 2002-2003. *Journal of the American Medical Association, 290*, 1190-1196.

Beale, A. V. (1970). The evolution of college admission requirements. *National ACAC Journal, 15*, 14-15.

Benbow, C. P. & Lubinski, D. (1993). Psychological profiles of the mathematical talented: Some sex differences and evidence supporting their biological basis. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), Ciba Foundation Symposium: Vol. 174. *Experimental and theoretical studies of consciousness* (pp. 44-66). Chichester, NY: Wiley.

Berger, A. G. (1961). How the college admission officer views counselor requirements. *Vocational Guidance Quarterly, 9*, 256.

The black-White scoring gap on SAT II Achievement Tests (2003). *Journal of Blacks in Higher Education, 38*, 61-63.

Borland, P. (Ed.) (1995) *Gender-Fair Math*. Newton, MA: WEEA Publishing Center.

Bowles, F. H. (1956). *The evolution of admission requirements: The interaction of school and college*. New York: College Entrance Examination Board.

Breland, H. M. (1979). *Population validity and college entrance measures* (College Board Rep. No. 79-8). New York: College Entrance Examination Board.

Bridgeman, B. & McHale, F. (1996). *Gender and ethnic group differences on the GMAT analytical writing assessment*. (ERIC Document Reproduction Service No. ED 400330).

Broome, E. C. (1903). *A historical and critical discussion of college admission requirements.* New York: Macmillan.

Bruschi, B. A., & Anderson, B. T. (1994, February). *Gender and ethnic differences in science achievement of nine, thirteen, and seventeen year old students*. Paper presented at the annual meeting of the Eastern Educational Research Association, Sarasota, FL.

Burton, N. W., Lewis, C., & Robertson, N. (1988). *Sex differences in SAT scores* (College Board Rep. No. 88-9). New York: College Entrance Examination Board.

Burton, N. W., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980* (College Board Rep. No. 01-2). New York: College Entrance Examination Board.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica, 28*, 591-605.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. *Journal of Educational Measurement, 5*, 115-124.

Cole, N. S. (1972). *Bias in selection* (ACT Research Rep. No. 51). Iowa City, IA: The American College Testing Program.

Colliver, J. A., Verhulst, S. J., & Williams, R. G. (1989). Using standardized patient examination to establish the predictive validity of the MCAT and undergraduate GPA as admissions criteria. *Academic Medicine, 64*, 482-484.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Harcourt Brace Jovanovich College.

Cronbach, L. (1971). Test validation. In R. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.

Cronbach, L., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 167-186.

Crouse, J., & Trusheim, D. (1988). *The case against the SAT*. Chicago: University of Chicago.

Cullen, T. J., Peckham, P. D, & Schwarz, M. R. (1980). Predicting first-quarter test scores from the new Medical College Admission Test. *Journal of Medical Education, 55*, 393-398.

Davidson, R., & MacKinnon, J. G. (1993). Estimation and inference in Econometrics.

New York: Oxford University Press.

Dollinger, S. J. (1989). Predictive validity of the Graduate Record Examination in Clinical Psychology. *Professional Psychology: Research and Practice, 20*, 56-58.

Ebel, R. L. (1978, October). *The role of standardized tests in college admissions, credit, and placement*. Paper presented at the National Forum of College Entrance Examination Admission, New York.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.) Englewood Cliffs, NJ: Prentice Hall.

Edwards, A. L. (1960). Experimental design in psychological research (Revised ed.). New York: Holt, Rinehart, and Winston.

Elam, C. L., Burke, M. M, Wiggs, J. S., & Speck, D. F. (1998). The medical school interview: Perspectives on preparation. *NACADA Journal, 18*, 28-32.

Elam, C. L., & Johnson, M. M. (1997). An analysis of admissions committees voting patterns. *Academic medicine, 72* (Suppl.), S45-S48.

Elam, C. L., Seaver, D. C., Berres, P. N., & Brandt, B. F. (2000). An overview of the admission processes for medical, dental, pharmacy, physical therapy, and physician assistant programs, *NACADA Journal, 20*, 24-31.

Elam, C. L., Seaver, D. C., Berres, P. N., & Brandt, B. F. (2002). Preparation for medical, dental, pharmacy, physical therapy, and physician assistant careers: Helping students gain a competitive edge. *Journal of College Admission, 176*, 16-21.

Elam, C. L., Stratton, T. D., & Lieber, K. L. (2002). Review, deliberation, and voting: A study of selection decisions in a medical school admissions committee. Teaching and Learning in Medicine, 14, 98-103.

Elam, C. L., Taylor, E. D., & Strother, E. N. (1996). Preparing for medical school and the medical profession: Advice to advisors. *NACADA Journal, 16*, 34-38.

Elliott, R., & Strentra, A. C. (1988). Effects of improving the reliability of the GPA on prediction generally and on comparative predictions for gender and race particularly. *Journal of Educational Measurement, 25*, 333-347.

Erdmann, J. B. (1977). *New Medical College Admission Test: Interpretive manual*. Washington, DC: Association of American Medical Colleges.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105-146). Phoenix, AZ: Oryx Press.

Fennema, E., & Carpenter, T. (1981). The second national assessment and sex-related differences in mathematics. *Mathematics Teacher, 74*, 554-559.

Ferguson, R.L. (1975, April). The *use of assessment in the selection of college students in the United States*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Fincher, C. (1992, September). *Admission policies and practices in U.S. colleges and universities*. Paper presented at the annual conference of the International Association for Educational Assessment, Dublin, Ireland.

Fine, B. (1946). *Admission to American colleges: A study of current policy and practice*. New York: Harper.

Fishman, J. A., & Pasanella, A. K. (1960). College admissions-selection studies. *Review of Educational Research, 30*, 298.

Fleming, J., & Garcia, N. (1998). Are standardized test scores fair to African Americans?

Predictive validity of the SAT in Black and White Institutions. *Journal of Higher Education, 69*, 471-495.

Fox, J. (1997). Applied regression analysis, linear models, and related methods. Thousand Oaks, CA: Sage.

Friedman, C. P., & Bakewell, W. E. (1980). Incremental validity of the new MCAT. *Journal of Medical Education, 55*, 399-404.

Gilbert, G. E., Basco, G. E., Blue, W. T., & O'Sullivan, P. S. (2002). Predictive validity of the Medical College Admissions Test Writing Sample for United States Medical Licensing Examination Steps 1 and 2. *Advances in Health Sciences Education, 7*, 191-200.

Givner, N., & Hynes, K. (1979). Achievement test validity: Correcting for restriction effects. *College and University, 54*, 119-123.

Glaser, K., Hojat, M., Veloski, J. J., Blacklow, R. S., & Goepp, C. E. (1992). Science, verbal, or quantitative skills: Which is the most important predictor of physician competence? *Educational and Psychological Measurement, 52*, 395-406.

Gottheil, E., & Michael, C. M. (1957). Predictor variables employed in research on selection of medical students. *Journal of Medical Education, 32*, 144-147.

Gough, H. G., Hall, W. B, & Harris, R. E. (1963). Admission procedures as forecasters of performance in medical training. *Journal of Medical Education, 38*, 983-998.

Grandy, J. (1994). *Trends and Profiles: Statistics about general test examinees by sex and ethnicity* (GRE Professional Rep. No. RR-94-1). Princeton NJ: Educational Testing Service.

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.

Gulliksen, H., & Wilks, S. S. (1950). Regression tests for several samples. *Psychometrika, 15*, 91-114.

Halpern, D. F. (1992). *Sex differences in cognitive abilities* (2nd ed.). Hillsdale, NJ: Erlbaum.

Halpern, D. F., Haviland, M. G., & Killian, C. D. (1998). Handedness and sex differences in intelligence: Evidence from the Medical College Admission Test. *Brain and Cognition, 38*, 87-101.

Hanlon, L. W. (1964). College grades and admission to medical schools. *Journal of Higher Education, 35*, 93-96.

Harris, M. B. (1998). Basic statistics for behavioral science research (2nd ed.). Needham Heights, MA: Allyn and Bacon.

Hill, J. K. (1959). Assessment of intellectual promise for medical school. *Journal of Medical Education, 34*, 959-964.

Hirschberg, N. W. (1977). Predicting performance in graduate school. In M.F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes in applied settings* (pp. 95-123). New York: Academic Press.

Hojat, M., Erdmann, J. B., Veloski, J. J., Nasca, T. J., Callahan, C. A., Julian, E., et al. (2000). A validity study of the Writing Sample section of the Medical College Admission Test. *Academic Medicine, 75* (Suppl.), S25-S27.

Holden, C. (1989, March 17). MCAT to stress thinking, writing. *Science, 243*, 1431.

House, J. D. (1998, May). *Gender differences in prediction of graduate course* *performance from admissions test scores: An empirical example of statistical methods for investigating prediction bias*. Paper presented at the annual forum of the Association for Institutional Research, Minneapolis, MN.

House, J. D., Gupta, S., & Xiao, B. (1997, November). *Gender differences in prediction of grade performance from Graduate Record Examination scores*. Paper presented at annual meeting of the Illinois Association for Institutional Research, Chicago.

House, J. D., & Keeley, E. J. (1993, November). *Differential prediction of graduate student achievement from Miller Analogies Test scores*. Paper presented at the annual meeting of the Illinois Association for Institutional Research, Oakbrook Terrace, IL.

Howell, D. C. (1995). *Fundamental statistics for the behavioral sciences* (3rd ed.). Belmont, CA: Duxbury.

Humphreys, L. G. (1962). The nature and organization of human abilities. *The Yearbook of the National Council on Measurement in Education, 19,* 16-19.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender difference in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139-155.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*, 53-69.

Iserson, K. V. (1997). *Get into medical school: A guide for the perplexed*. Tucson, AZ: Galen Press.

Jencks, C., & Phillips, M. (Eds.) (1998). *The Black-White Test Score Gap.* Washington, DC: Brookings Institution.

Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.

Johnson, D. G. (1983). *Physicians in the making*. San Francisco: Jossey-Bass.

Johnson, D. G., Lloyd, S. M., Jones, R. F, & Anderson, J. (1986). Predicting performance at a predominantly black medical school. *Academic Medicine, 61*, 629-639.

Johnson, E. K., & Edwards, J. C. (1991). Current practices in admissions interviews at U.S. medical schools. *Academic Medicine, 66*, 408-412.

Jones, R. F., & Mitchell, K. (1986, April). *Racial/ethnic differences in the predictive validity of MCAT scores*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Jones, R. F., & Thomae-Forgues, M. (1984). Validity of the MCAT for predicting performance in the first two years of medical school. *Journal of Medical Education, 59*, 455-464.

Jones, R. F., & Vanyur, S. (1985, April). *Gender-related test bias for the Medical College Admission Test.* Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.

Julian, E., & Lockwood, J. (2000). Predictive validity of the medical college admission test. *Contemporary Issues in Medical Education, 3*, 2.

Juola, A. E. (1968). Illustrative problems in college-level grading. *Personal and Guidance Journal, 47*, 29-33.

Kelley, T. L. (1947). *Fundamentals of statistics*. Cambridge, MA: Harvard University.

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

Kingsley, C. D. (1913). *College entrance requirements*. Washington, DC: Government Printing Office.

Koenig, J., Huff, K., & Julian, E. (2002). *Predictive validity of the Medical College Admission Test* (MCAT Monograph No. 8). Washington, DC: Association of American Medical Colleges.

Koenig, J. A., Sireci, S. G., & Wiley, A. (1998). Evaluating the predictive validity of MCAT scores across diverse applicant groups. *Academic Medicine, 73*, 1095-1106.

Kuncel, N. R., Campbell, J. P., & Ones, D. S. (1998). Validity of the Graduate Record Examination: Estimated or tacitly known. *American Psychologist, 53*, 567-568.

Kuncel, N. R., Hezlett, S. A. & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*, 162-181.

Lakhan, S. E. (2003). Diversification of U.S. medical schools via affirmative action implementation. *BMC Medical Education, 3*, 1-6.

Lee, R., Miller, K. J., & Graham, W. K. (1982). Correction for restriction of range and attenuation in criterion-related validation studies. *Journal of Applied Psychology, 67*, 637-639.

Linn, R. L. (1973). Fair test use in selection. *Review of Educational Research, 43*, 139-161.

Linn, R. L. (1978). Single-group validity, differential validity, and differential prediction. Journal of Applied Psychology, 63, 507-512.

Linn, R. L. (1982). Ability testing: Individual differences, prediction, and differential prediction. In A. Wigor & W. Garner (Eds.), *Ability testing: Uses, consequences, and controversies* (pp. 335-338). Washington, DC: National Academy Press.

Linn, R. L. (1984). Selection bias: Multiple meanings. *Journal of Educational Measurement, 21*, 33-48.

Linn, R. L. (1990). Admissions testing: Recommended uses, validity, differential prediction, and coaching. *Applied Measurement in Education, 34*, 297-318.

Linn, R. L. (1993). The use of differential item functioning statistics: A discussion of current practices and future implications. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 349-364). Hillsdale, NJ: Lawrence Erlbaum.

Linn, R. L. (2001). A century of standardized testing: Controversies and pendulum swings. *Educational Assessment, 7*, 29-38.

Linn, R. L., & Hastings, C. N. (1984). Group differentiated prediction. *Applied Psychological Measurement, 8*, 165-172.

Littlemeyer, M. H., & Mauney, A. C. (1977). *New MCAT student manual* (2nd ed.). Washington, DC: Association of American Medical Colleges.

Lord, F. M. & Novick, M. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Ludmeyer, K. M. (1999). *Time to heal: American Medical education from the turn of the century to the managed care era*. New York: Oxford University.

Maccoby, E. E. & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University.

Manning, W. H., Willingham, W. W., & Breland, H. M. (1977). *Selective admissions in higher education: Comments and Recommendation of the Carnegie Council on policy studies in higher education*. San Francisco: Jossey-Bass.

Matheson, D. (2001). *Enquiry about the Chow Test in SPSS*. Retrieved July 7, 2004, from http://pages.infinit.net/rlevesqu/Syntax/RegressionRepeatedMeasure/ChowTest.txt

McGaghie, W. C. (1987). Liberal education and medical school admission. *Journal of General Internal Medicine, 2*, 361-363.

McGaghie, W. C. (2002a). Assessing readiness for medical education: Evolution of the Medical College Admission Test. *Journal of American Medical Association, 288*, 1085-1090.

McGaghie, W. C. (2002b). Student selection. In G. R. Norman, C. P. M. van der Vleuten & D. I. Newble (Eds.), *International handbook of research on medical education: Part one* (pp. 97- 126). Norwell, MA: Kluwer Academic.

McGuire, F. L. (1980). The new MCAT and medical student performance. *Journal of Medical Education, 55*, 405-408.

Medical College Admission Test. (2003). *MCAT student manual*. Washington, DC: Association of American Medical Colleges.

Mehrens, W. A., & Lehmann, I. J. (1987). *Using standardized tests in education* (4th ed.). New York: Longman.

Mertler, C. A., & Vannatta, R. A. (2002). *Advanced multivariate statistical methods: Practical application and interpretation* (2nd ed.). Los Angeles: Pyrczak.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Phoenix, AZ: Oryx Press.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

Messick, S. (2000). Consequences of test interpretation and use: The fusion of validity and values in psychological assessment. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment* (pp. 3-20). Norwell, MA: Kluwer Academic.

Meyer, M. R. (1989). Gender differences in mathematics. In M. M. Lindquist (Ed.), *Results from the fourth mathematics assessment* (pp. 149-159). Reston, VA: National Council of Teachers of Mathematics.

Mitchell, K. J. (1987). Use of MCAT data in selecting students for admission to medical school. *Academic Medicine, 62*, 871-879.

Mitchell, K. J. (1990). Traditional predictors of performance in medical school. *Academic Medicine, 65*, 149-158.

Mitchell, K. J. (1991). *Use of MCAT data in admissions: A guide for medical school admissions officers and faculty*. Washington, DC: Association of American Medical Colleges.

Mitchell, K. J., Haynes, R., & Koenig, J. (1994). Assessing the validity of the updated Medical College Admission Test. *Academic Medicine, 69*, 394-401.

Moss, F. A. (1930). Scholastic aptitude tests for medical students. *Journal of Association of American Medical College, 5*, 90-110.

Moss, F. A. (1938). Report of the committee on aptitude tests for medical students. *Journal of Association of American Medical Colleges, 13*, 177-189.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement, 56*, 63-75.

Mulvenon, S. W., Stegman, C., Thorn, A., & Thomas, S. (1999). Selection for college admission: Refining traditional models. *Journal of College Admission, 162*, 20-27.

Nettles, A. L., & Nettles, M. T. (Eds.) (1999). *Measuring up: Challenges minorities face in educational assessment*. Norwell, MA: Kluwer Academic.

Norusis, M. J. (1998). SPSS 8.0: A guide to data analysis. Upper Saddle River, NJ: Prentice Hall.

Nowacek, G. A., Bailey, B. A., & Sturgill, B. C. (1996). Influence of the interview on the evaluation of applicants to medical school. *Academic Medicine, 71*, 1093-1095.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Olmstead, A. L., & Sheffrin, S. M. (1981). The medical school admission process: An empirical investigation. *Journal of Human Resources, 16*, 459-467.

Pearson, K. (1903). Mathematical contributions to the theory of evolution-XI. On influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society, London, Series A, 200*, 1-66.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and* Prediction (2nd ed.). New York: Holt, Reinhart, and Winston.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Fort Worth, TX: Harcourt Brace College.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

The persisting racial score gap on graduate and professional school admission tests (2003). *Journal of Blacks in Higher Education, 38*, 16-20.

Powis, D. A., McManus, I. C., & Cleve-Hogg, D. (1992). Selection of medical students: Philosophic, political, social, and educational bases. *Teaching and Learning Medicine, 4*, 25-34.

Raju, N. S., & Brand, P. A. (2003). Determining the significance of correlations corrected for unreliability and range restriction. *Applied Psychological Measurement, 27*, 52-71.

Ramist, L. (1984). Predictive validity of ATP tests. In T.F. Donlon (Ed.), *College Board technical handbook for the Scholastic Aptitude and Achievement Tests* (pp. 141-170). New York: College Entrance Examination Board.

Ramos, S., Croen, L., & Haddow, S. (1986). Predictors of preclinical and clinical performance of minority medical students. *Journal of the National Medical Association, 78*, 601-607.

Reynolds, C. R. (1982). Methods for detecting construct and predictive bias. In R. A. Berk (Ed.), *Handbook of methods for detecting test bias* (pp.199-227). Baltimore: Johns Hopkins University.

Rudner, L. M. (1994). *Questions to ask when evaluating tests*. Washington, DC: ERIC

Clearinghouse on Assessment and Evaluation. (ERIC Document Reproduction Service No. ED 385607).

Sanazaro, P. J., & Hutchins, E. B. (1963). The origin and rationale of the Medical College Admission Test. *Journal of Medical Education, 38*, 1044-1050.

Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology, 61*, 473-485.

Sedlacek, W. E. (1967). *Medical College Admission Test: Handbook for admission committees* (2nd ed.). Evanston, IL: Association of American Medical Colleges.

Shea, J. A, & Fortna, G. S. (2002). Psychometric methods. In G. R. Norman, C. P. M. van der Vleuten, & D. I. Newble (Eds.), *International handbook of research on medical education: Part one* (pp. 97- 126). Norwell, MA: Kluwer Academic.

Shepard, A. (1987). The selection of medical students. *Journal of the American Medical Association, 257*, 2291-2292.

Shepard, L. A. (1987). Bias in tests of achievement and scholastic aptitude. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 177-190). Philadelphia: Falmer.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*, 72-101.

Stefanu, C., & Farmer, T. A. (1971). The differential predictive validity of Science MCAT in the admissions process. *Journal of Medical Education, 46*, 461-463.

Stevens, J. (1996). *Applied multivariate statistics* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Swanson, A. G., & Mitchell, K. J. (1989). MCAT responds to changes in medical education and physician practice. *Journal of the American Medical Association, 262*, 261-263.

Tabachnick, B. G., & Fidell, L. S. (2001). Using multivariate statistics (4th ed.). Boston: Allyn and Bacon.

Tatsuoka, M. M. (1969). *Selected topics in advanced statistics: An elementary approach.* Champaign, IL: The Institute for Personality and Ability Testing.

Thomas, C. L. (1972, April). *The relative effectiveness of high school grades and standardized test scores for predicting college grades for black students*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. New York: Wiley.

Thorndike, R. L. (1982). *Applied Psychometrics*. Boston: Houghton Mifflin.

Traxler, A. E., & Townsend, A. (1953). *Improving transition from school to college: How can school and college best cooperate*. New York: Harper.

Vancouver, J. B., Reinhart, M. A., Solomon, D. J., & Haf, J. J. (1990). Testing for validity and bias in the use of GPA and the MCAT in the selection of medical school students. *Academic Medicine, 65*, 694-697.

Veloski, J. J., Callahan, C. A., Xu, G., Hojat, M., & Nash, D. B. (2000). Prediction of students' performance on licensing examinations using age, race, sex, undergraduate GPA, and MCAT scores. *Academic Medicine, 75* (Suppl.), S28-S30.

Wainer, H., & Braun, H. I. (Eds.) (1988). *Test validity*. Hillsdale, NJ: Erlbaum.

Westoff, L. A. (1980). A new era in admissions. *Focus, 7*, 1-24.

Wigdor, A. K., & Garner, W. R. (Eds.) (1982). *Ability testing: Uses, consequences, and controversies*. Washington, DC: National Academic.

Wightman, L. F. (2003). Standardized testing and equal access: A tutorial. In M. J. Chang (Ed.), *Compelling interest: Examining the evidence of racial dynamics in colleges and universities* (pp. 49-96). Stanford, CA: Stanford University.

Wiley, A., & Koenig, J. A. (1996). The validity of the Medical College Admission Test for predicting performance in the first two years of medical school. *Academic Medicine, 71* (Suppl.), S83-S85.

Wilder, G., & Powell, K. (1989). *Sex differences in test performance: A survey of the literature* (College Board Rep. No. 89-3). New York: College Entrance Examination Board.

Wilkinson, E. P. (2004). *Medical school curricula*. Retrieved April 7, 2004, from http://www.faqs.org/faqs/medicine/education-faq/part2/section-2.html.

Willingham, W. W. (1974, January 25). Predicting success in graduate education. *Science, 183*, 273-278.

Willingham, W. W. & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum.

Willingham, W. W., Lewis, C., Morgan, R., & Ramist, L. (1990). *Predicting college grades: An analysis of institutional trends over two decades*. Princeton, NJ: Education Testing Service.

Wilson, K. M. (1983). *A review of research on the prediction of academic performance* *after the freshman year* (College Board Rep. No. 83-2). New York: College Entrance Examination Board.

Young, J. W. (1991). Improving the prediction of college performance of ethnic minorities using IRT-based GPA. *Applied Measurement in Education, 4*, 229-239.

Young, J. W. (1994). Differential prediction of college grades by gender and by ethnicity: A replicated study. *Educational and Psychological Measurement, 54*, 1022-1029.

Young, J. W. (2001). *Differential validity, differential prediction, and college admission testing: A comprehensive review and analysis* (College Board Rep. No. 01-6). New York: College Entrance Examination Board.

Young, J. W., & Fisler, J. L. (2000). Sex differences on the SAT: An analysis of demographic and educational variables. *Research in Higher Education, 41*, 401-416.

Zar, J. (1999). Biostatistical Analysis (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Appendixes

Appendix A

MCAT Confidential Certification Statement

ASSOCIATION OF AMERICAN MEDICAL COLLEGES
2450 N STREET, NW, WASHINGTON, DC 20037-1127
PHONE 202-828-0400 FAX 202-828-1125
HTTP://WWW.AAMC.ORG

Medical College Admission Test (MCAT)
Confidentiality Certification Statement

I recognize that the MCAT is a worldwide secure examination and that it is necessary to prevent dissemination of information regarding the MCAT and to prevent those who have scored the MCAT from assisting persons preparing to take the exam. As a condition of using MCAT data, I agree to adhere to the Association of American Medical College's (AAMC) security requirements which include the following:

I certify that I have not had any connection to (e.g., been employed by) an MCAT review course or an MCAT preparation book publisher in the last five years.

I do not plan to take the MCAT within the next two years, and I agree that I will be prohibited from doing so.

I agree not to disclose information about the structure, emphasis, composition, or scoring of the MCAT to anyone other than my designated academic advisors or the AAMC.

I acknowledge AAMC's exclusive property rights in the MCAT materials.

I agree not to directly or indirectly use the information gained from serving on the MCAT project in the preparation of any public or private coursework for the MCAT including MCAT review courses, workshops, individual or group tutoring activities, or any other MCAT test preparation activity.

I agree to share the results of my research using MCAT data with the Association of American Medical Colleges.

I agree that AAMC has that right to prevent or stop me from doing any of the activities noted above.

Signature Lydia Kyei-Blankson ________ Date 1-26-04
Lydia Kyei-Blankson Social Security Number:

Signature ________________ Date 3-4-04
Association of American Medical Colleges

Appendix B

Exemption Letter from the Ohio University Institutional Review Board

OHIO UNIVERSITY

Office of the Vice President
for Research

04E045

Office of Research Compliance
Research and Technology
Center 117
Athens OH 45701-2979

T: 740.593.0664
F: 740.593.9838
www.ohiou.edu/research

A determination has been made that the following research study is exempt from IRB review because it involves:

Category 4 - research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens if publicly available or recorded without identifiers

Project Title: Predictive Validity, Differential Validity, and Differential Prediction of the Subtests of the Medical College Admission Test

Project Director: Lydia Kyei-Blankson

Department: Educational Studies

Advisor: George Johanson

Appendix C

Predictor and Criterion Means and Standard Deviations by Gender and Ethnicity

Table 21

Variable Means and Standard Deviations by Gender for School 1

	Men (n = 147)		Women (n = 62)	
Variable	M	SD	M	SD
BS	9.60	1.62	8.73	1.75
PS	9.78	1.90	8.45	1.88
VR	9.85	1.66	9.45	1.64
WS	6.40	1.99	6.58	1.97
NGPA	3.35	.39	3.41	.35
SGPA	3.25	.49	3.17	.50
MGPA	2.80	.65	2.73	.70

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 22

Variable Means and Standard Deviations by Ethnicity for School 1

	White (n = 157)		Black (n = 30)		Asian (n = 13)		Hispanic (n = 4)	
Variable	M	SD	M	SD	M	SD	M	SD
BS	9.64	1.46	7.27	1.39	10.15	1.82	8.20	1.83
PS	9.75	1.71	7.00	1.62	10.23	1.59	7.82	1.91
VR	10.14	1.40	7.77	1.79	9.46	1.27	8.19	1.95
WS	6.63	1.99	5.73	1.62	5.92	2.06	5.81	1.97
NGPA	3.41	.37	3.08	.41	3.55	.24	3.33	.25
SGPA	3.33	.41	2.65	.59	3.42	.35	2.86	.32
MGPA	2.83	.64	2.63	.79	2.56	.65	2.31	.69

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 23

Variable Means and Standard Deviations by Gender for School 2

	Men (n = 74)		Women (n = 62)	
Variable	M	SD	M	SD
BS	11.00	1.97	10.26	1.68
PS	10.88	2.05	9.87	1.93
VR	9.50	1.89	9.48	1.91
WS	6.46	1.99	6.21	1.81
NGPA	3.65	.26	3.61	.29
SGPA	3.56	.38	3.55	.35
MGPA	8.12	1.01	7.72	1.16

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 24

Variable Means and Standard Deviations by Ethnicity for School 2

	White (n = 57)		Black (n = 10)		Asian (n = 43)		Hispanic (n = 19)	
Variable	M	SD	M	SD	M	SD	M	SD
BS	11.14	1.63	9.20	1.87	10.91	1.65	9.37	2.03
PS	10.82	1.71	8.20	2.39	11.12	1.68	8.89	1.97
VR	10.30	1.45	7.70	2.16	9.42	1.91	8.21	1.51
WS	6.42	2.14	6.40	2.63	6.58	1.33	5.11	1.56
NGPA	3.71	.23	3.28	.33	3.71	.17	3.43	.34
SGPA	3.65	.23	2.87	.49	3.70	.26	3.30	.39
MGPA	8.45	.64	6.08	1.41	8.10	.77	7.32	.98

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 25

Variable Means and Standard Deviations by Gender for School 3

	Men (n = 120)		Women (n = 73)	
Variable	M	SD	M	SD
BS	8.88	1.60	8.47	1.73
PS	8.74	1.68	7.88	1.47
VR	8.76	1.80	8.85	2.15
WS	6.00	1.84	6.62	1.71
NGPA	3.51	.38	3.64	.31
SGPA	3.38	.46	3.41	.45
MGPA	83.93	6.19	83.10	5.91

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 26

Variable Means and Standard Deviations by Ethnicity for School 3

	White (n = 158)		Black (n = 3)		Asian (n = 19)		Hispanic (n = 9)	
Variable	M	SD	M	SD	M	SD	M	SD
BS	8.87	1.55	7.00	1.00	8.16	1.92	7.44	2.01
PS	8.54	1.67	8.00	1.00	8.00	1.56	7.11	1.36
VR	8.95	1.80	9.00	1.00	7.79	2.35	7.44	2.51
WS	6.34	1.75	5.00	2.00	5.79	1.93	5.89	2.52
NGPA	3.62	.30	2.59	.18	3.34	.46	3.24	.41
SGPA	3.49	.32	2.10	.15	3.15	.65	2.68	.60
MGPA	84.46	5.46	74.02	1.86	81.54	4.95	75.33	9.55

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 27

Variable Means and Standard Deviations by Gender for School 4

	Men (n = 325)		Women (n = 195)	
Variable	M	SD	M	SD
BS	9.50	2.02	8.20	2.06
PS	9.62	2.38	7.98	2.14
VR	9.09	2.13	8.79	1.88
WS	5.91	1.94	6.15	2.01
NGPA	3.44	.37	3.45	.38
SGPA	3.22	.46	3.02	.58
MGPA	70.27	9.03	67.02	10.62

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 28

Variable Means and Standard Deviations by Ethnicity for School 4

	White (n = 268)		Black (n = 87)		Asian (n = 96)		Hispanic (n = 54)	
Variable	M	SD	M	SD	M	SD	M	SD
BS	9.62	1.77	6.54	1.45	10.31	1.52	7.52	1.55
PS	9.75	2.12	6.37	1.09	10.42	1.87	7.09	1.84
VR	9.77	1.59	6.90	1.82	9.60	1.37	7.17	2.05
WS	6.15	1.96	5.46	2.02	6.25	1.83	5.67	2.05
NGPA	3.55	.33	3.17	.37	3.55	.30	3.15	.37
SGPA	3.33	.40	2.60	.51	3.33	.37	2.73	.44
MGPA	72.34	8.48	60.04	8.18	72.09	8.00	63.67	9.67

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 29

Variable Means and Standard Deviations by Gender for School 5

	Men (n = 183)		Women (n = 108)	
Variable	M	SD	M	SD
BS	9.02	1.44	8.41	1.46
PS	9.04	1.74	7.83	1.62
VR	9.11	1.59	9.63	1.52
WS	5.96	2.09	6.03	2.04
NGPA	3.62	.29	3.66	.32
SGPA	3.40	.37	3.43	.39
MGPA	3.11	.56	2.94	.55

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 30

Variable Means and Standard Deviations by Ethnicity for School 5

	White (n = 234)		Black (n = 7)		Asian (n = 18)		Hispanic (n = 5)	
Variable	M	SD	M	SD	M	SD	M	SD
BS	8.91	1.41	7.00	1.16	9.11	1.68	8.80	1.30
PS	8.74	1.78	7.14	1.57	8.56	1.82	8.60	2.07
VR	9.53	1.44	6.71	1.60	8.39	1.82	9.20	1.79
WS	6.13	2.03	3.86	1.46	5.39	2.15	7.20	1.30
NGPA	3.66	.29	3.22	.32	3.63	.29	3.61	.36
SGPA	3.45	.36	2.85	.42	3.39	.44	3.16	.60
MGPA	3.08	.56	2.40	.60	3.08	.37	3.08	.13

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA

Table 31

Variable Means and Standard Deviations by Gender for School 6

	Men (n = 231)		Women (n = 141)	
Variable	M	SD	M	SD
BS	9.43	1.42	8.89	1.62
PS	9.48	1.69	8.24	1.50
VR	9.37	1.47	9.53	1.54
WS	6.10	2.05	6.33	1.85
NGPA	3.53	.36	3.59	.28
SGPA	3.37	.41	3.32	.41
MGPA	3.12	.75	2.94	.72

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 32

Variable Means and Standard Deviations by Ethnicity for School 6

	White (n = 241)		Black (n = 11)		Asian (n = 56)		Hispanic (n = 58)	
Variable	M	SD	M	SD	M	SD	M	SD
BS	9.48	1.38	6.09	1.14	9.57	1.13	8.36	1.54
PS	9.22	1.60	6.27	1.35	9.64	1.53	7.97	1.68
VR	9.72	1.39	7.55	1.92	9.18	1.40	8.88	1.53
WS	6.27	2.00	4.91	1.45	6.25	2.13	5.98	1.84
NGPA	3.56	.33	3.45	.40	3.67	.27	3.46	.35
SGPA	3.41	.37	2.97	.50	3.39	.39	3.16	.46
MGPA	3.21	.69	2.12	.47	3.03	.67	2.62	.76

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 33

Variable Means and Standard Deviations by Gender for School 7

	Men (n = 155)		Women (n = 107)	
Variable	M	SD	M	SD
BS	9.65	1.82	8.21	1.74
PS	9.43	2.08	7.94	1.68
VR	8.90	1.82	8.44	1.77
WS	6.26	1.90	5.85	1.76
NGPA	3.42	.35	3.51	.30
SGPA	3.34	.40	3.27	.47
MGPA	2.62	.78	2.14	.73

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 34

Variable Means and Standard Deviations by Ethnicity for School 7

	White (n = 146)		Black (n = 46)		Asian (n = 40)		Hispanic (n = 23)	
Variable	M	SD	M	SD	M	SD	M	SD
BS	9.58	1.45	6.93	1.60	10.15	1.42	7.65	1.82
PS	9.38	1.75	6.57	1.12	10.05	1.72	7.22	1.65
VR	9.34	1.43	6.98	1.87	8.92	1.69	7.57	1.62
WS	6.30	1.84	5.15	1.67	6.70	1.84	5.48	1.65
NGPA	3.54	.30	3.23	.37	3.53	.23	3.29	.34
SGPA	3.46	.28	2.83	.45	3.51	.30	2.96	.44
MGPA	2.56	.68	1.72	.77	2.86	.65	1.98	76

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 35

Variable Means and Standard Deviations by Gender for School 8

	Men (n = 146)		Women (n = 110)	
Variable	M	SD	M	SD
BS	10.58	1.70	9.85	1.83
PS	10.86	2.00	9.61	177
VR	10.14	1.43	9.92	1.83
WS	6.84	2.12	7.08	2.03
NGPA	3.57	.30	3.60	.29
SGPA	3.55	.36	3.48	.39
MGPA	9.29	1.78	9.06	1.94

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 36

Variable Means and Standard Deviations by Ethnicity for School 8

	White (n = 181)		Black (n = 27)		Asian (n = 33)		Hispanic (n = 3)	
Variable	M	SD	M	SD	M	SD	M	SD
BS	10.46	1.58	8.15	1.58	10.67	1.51	12.33	1.16
PS	10.65	1.76	7.67	1.64	10.61	1.56	11.67	2.08
VR	10.42	1.16	8.00	2.24	9.79	1.19	11.67	.58
WS	7.04	2.08	6.11	1.85	7.39	1.98	8.00	1.73
NGPA	3.61	.27	3.31	.39	3.68	.17	3.66	.25
SGPA	3.59	.28	2.93	.46	3.63	.23	3.71	.17
MGPA	9.47	1.68	7.41	2.11	9.23	1.60	9.25	1.43

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 37

Variable Means and Standard Deviations by Gender for School 9

	Men (n = 117)		Women (n = 109)	
Variable	M	SD	M	SD
BS	11.72	1.60	11.06	1.66
PS	12.14	1.70	10.95	1.77
VR	10.68	1.46	10.82	1.29
WS	6.97	1.72	7.25	1.72
NGPA	3.74	.21	3.79	.17
SGPA	3.77	.22	3.75	.23
MGPA	2.61	.70	2.61	.71

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 38

Variable Means and Standard Deviations by Ethnicity for School 9

	White (n = 156)		Black (n = 16)		Asian (n = 48)		Hispanic (n = 1)	
Variable	M	SD	M	SD	M	SD	M	SD
BS	11.58	1.44	8.94	1.48	11.67	1.67	12.00	
PS	11.76	1.55	8.31	2.41	12.00	1.43	12.00	
VR	10.98	1.33	8.94	1.48	10.56	.99	11.00	
WS	7.16	1.70	6.31	1.85	7.21	1.73	8.00	
NGPA	3.77	.20	3.70	.21	3.77	.19	3.82	
SGPA	3.79	.21	3.45	.26	3.75	.20	3.93	
MGPA	2.67	.68	2.20	.81	2.61	.70	2.05	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 39

Variable Means and Standard Deviations by Gender for School 10

	Men (n = 86)		Women (n = 102)	
Variable	M	SD	M	SD
BS	9.72	1.44	9.20	1.54
PS	9.52	1.85	8.37	1.45
VR	9.37	1.57	9.44	1.53
WS	6.67	2.26	6.56	1.85
NGPA	3.58	.32	3.65	.30
SGPA	3.46	.40	3.50	.39
MGPA	83.23	5.99	83.58	5.08

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 40

Variable Means and Standard Deviations by Gender for School 11

	Men (n = 93)		Women (n = 80)	
Variable	M	SD	M	SD
BS	9.87	1.57	9.07	1.59
PS	10.45	1.82	9.14	1.75
VR	9.46	1.82	9.52	1.69
WS	6.82	1.89	7.06	2.10
NGPA	3.40	.34	3.53	.28
SGPA	3.39	.38	3.35	.41
MGPA	2.37	.71	2.27	.79

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 41

Variable Means and Standard Deviations by Ethnicity for School 11

	White (n = 108)		Black (n = 14)		Asian (n = 32)		Hispanic (n = 8)	
Variable	M	SD	M	SD	M	SD	M	SD
BS	9.70	1.42	7.14	1.10	9.84	1.51	8.38	1.77
PS	9.93	1.67	7.43	1.16	10.53	1.95	8.13	1.73
VR	9.76	1.64	7.50	1.29	9.16	1.89	9.25	1.28
WS	7.28	1.90	5.71	2.09	6.59	2.00	6.00	2.33
NGPA	3.47	.33	3.33	.34	3.58	.20	3.16	.21
SGPA	3.44	.33	2.92	.47	3.47	.33	3.03	.52
MGPA	2.40	.71	1.36	.75	2.47	.62	1.59	.38

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 42

Variable Means and Standard Deviations by Gender for School 12

	Men (n = 65)		Women (n = 67)	
Variable	M	SD	M	SD
BS	9.06	1.52	8.45	1.43
PS	8.62	1.34	7.76	1.29
VR	8.88	1.54	8.91	1.61
WS	6.03	1.79	6.08	2.02
NGPA	3.61	.32	3.55	.35
SGPA	3.45	.35	3.25	.51
MGPA	2.98	.60	2.89	.30

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 43

Variable Means and Standard Deviations by Ethnicity for School 12

	White (n = 107)		Black (n = 12)		Asian (n = 8)		Hispanic (n = 1)	
Variable	M	SD	M	SD	M	SD	M	SD
BS	8.98	1.24	6.50	2.07	8.50	1.20	11.00	
PS	8.31	1.31	6.75	1.60	8.38	.74	10.00	
VR	9.24	1.25	6.42	1.44	8.38	1.85	11.00	
WS	6.27	1.90	5.33	1.72	4.63	.92	9.00	
NGPA	3.63	.28	3.04	.35	3.72	.23	3.63	
SGPA	3.43	.38	2.60	.35	3.57	.26	2.95	
MGPA	3.00	.45	2.59	.44	2.75	.46	3.00	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 44

Variable Means and Standard Deviations by Gender for School 13

	Men (n = 92)		Women (n = 48)	
Variable	M	SD	M	SD
BS	9.01	1.27	8.38	1.70
PS	8.59	1.56	7.79	1.52
VR	9.24	1.43	8.94	1.71
WS	5.72	2.10	6.21	1.92
NGPA	3.33	.38	3.56	.32
SGPA	3.28	.34	3.43	.36
MGPA	2.92	.66	2.97	.66

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 45

Variable Means and Standard Deviations by Ethnicity for School 13

	White (n = 117)		Black (n = 7)		Asian (n = 10)		Hispanic (n = 3)	
Variable	M	SD	M	SD	M	SD	M	SD
BS	8.86	1.33	7.14	1.68	9.20	1.99	8.67	2.52
PS	8.38	1.58	7.43	1.27	7.90	1.91	9.33	1.53
VR	9.22	1.53	8.29	1.80	8.70	1.25	9.33	2.08
WS	5.86	2.01	5.71	1.25	6.10	2.73	5.66	4.04
NGPA	3.39	.37	3.44	.45	3.64	.27	3.22	.77
SGPA	3.34	.34	3.17	.45	3.37	.22	3.49	.54
MGPA	2.99	.63	2.84	.38	2.34	.93	3.21	.25

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 46

Variable Means and Standard Deviations by Gender for School 14

	Men (n = 41)		Women (n = 48)	
Variable	M	SD	M	SD
BS	7.29	1.44	6.98	1.56
PS	7.34	1.65	6.98	1.18
VR	7.29	2.27	7.92	1.84
WS	4.83	1.79	5.50	2.05
NGPA	3.28	.36	3.38	.38
SGPA	2.92	.43	2.89	.56
MGPA	77.34	7.49	78.12	7.84

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA = Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 47

Variable Means and Standard Deviations by Ethnicity for School 14

	White (n = 6)		Black (n = 68)		Asian (n = 11)		Hispanic (n = 3)	
Variable	M	SD	M	SD	M	SD	M	SD
BS	7.83	1.94	6.84	1.37	8.18	1.54	7.33	1.16
PS	7.50	.55	6.82	1.30	8.55	1.37	8.33	2.08
VR	9.33	2.07	7.32	2.04	8.27	1.95	8.00	1.00
WS	5.67	2.25	5.12	2.14	5.18	2.14	5.33	1.53
NGPA	3.29	.24	3.32	.39	3.49	.36	3.11	.21
SGPA	3.11	.55	2.86	.52	3.09	.38	2.87	.52
MGPA	81.86	8.41	76.50	7.61	83.22	5.17	76.54	6.78

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Appendix D

Predictors and Criterion Intercorrelations by School and by Gender and Ethnicity

Table 48

Intercorrelations among Variables for School 1

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.644	.421	.054	.189	.341	.202
PS			.438	.055	.128	.376	.224
VR				.257	.217	.218	.214
WS					.264	.164	.109
NGPA						.714	.200
SGPA							.365

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 49

Intercorrelations among Variables by Gender for School 1

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
	Upper Triangle = Men[a] ; Lower Triangle = Women[b]						
BS		.628	.416	.108	.130	.274	.197
PS	.597		.395	.078	.081	.390	.250
VR	.394	.508		.328	.227	.233	.237
WS	-.029	.052	.105		.252	.192	.181
NGPA	.418	.365	.230	.289		.690	.158
SGPA	.464	.330	.161	.110	.813		.298
MGPA	.193	.154	.153	-.049	.314	.497	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 147. [b] n = 62.

Table 50

Intercorrelations among Variables by Ethnicity for School 1

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Whites[a]; Lower Triangle = Blacks[b]							
BS		.520	.227	-.078	.056	.111	.212
PS	.292		.241	.027	.008	.170	.283
VR	.275	.309		.275	.179	.036	.239
WS	.171	-.198	-.093		.246	.119	.155
NGPA	.078	-.166	-.179	.205		.678	.243
SGPA	.044	.107	-.182	.070	.747		.396
MGPA	.171	.019	-.042	-.328	.040	.302	
Upper Triangle = Asians[c] ; Lower Triangle = Hispanics[d]							
BS		.794	.220	.003	-.402	.220	.084
PS	.662		.440	-.121	-.337	.006	-.117
VR	.816	.324		-.049	-.253	-.424	.159
WS	.787	.072	.741		.356	-.013	.115
NGPA	.565	.417	.033	.532		.630	.355
SGPA	.662	.970	.215	.131	.613		.427
MGPA	-.486[c]	-.723	.103	-.165	-.874	-.868	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 157. [b] n = 30. [c] n = 13. [d] n = 4.

Table 51

Intercorrelations among Variables for School 2

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.762	.523	.261	.397	.427	.625
PS			.493	.279	.392	.469	.613
VR				.277	.354	.258	.573
WS					.153	.094	.171
NGPA						.676	.472
SGPA							.520

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 52

Intercorrelations among Variables by Gender for School 2

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
	Upper Triangle = Men[a] ; Lower Triangle = Women[b]						
BS		.773	.483	.252	.420	.463	.547
PS	.719		.533	.195	.453	.495	.575
VR	.606	.477		.258	.354	.275	.516
WS	.257	.378	.302		.085	.056	.137
NGPA	.364	.316	.357	.224		.725	.501
SGPA	.384	.455	.236	.143	.629		.600
MGPA	.691	.628	.649	.183	.426	.447	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 74. [b] n = 62.

Table 53

Intercorrelations among Variables by Ethnicity for School 2

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Whites[a]; Lower Triangle = Blacks[b]							
BS		.706	.435	.254	.325	.349	.393
PS	.857		.352	.299	.173	.325	.332
VR	.729	.421		.155	.245	.044	.152
WS	.613	.391	.511		.101	-.007	.182
NGPA	.845	.726	.538	.565		.422	.402
SGPA	.727	.680	.367	.461	.777		.473
MGPA	.540	.172	.656	.313	.387	-.226	
Upper Triangle = Asians[c] ; Lower Triangle = Hispanics[d]							
BS		.651	.370	.058	-.245	.146	.419
PS	.787		.371	.150	-.281	-.047	.327
VR	.281	.362		.249	-.103	-.076	.133
WS	.092	.094	.461		-.140	-.136	-.522
NGPA	.332	.321	.125	.037		.410	.160
SGPA	.062	.196	-.159	-.156	.702		.324
MGPA	.428	.532	.289	.303	-.184	-.191	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 57. [b] n = 10. [c] n = 43. [d] n = 19.

Table 54

Intercorrelations among Variables for School 3

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.581	.381	.114	.219	.349	.322
PS			.320	.067	.161	.313	.332
VR				.181	.315	.228	.255
WS					.257	.177	.117
NGPA						.688	.308
SGPA							.432

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 55

Intercorrelations among Variables by Gender for School 3

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Men[a] ; Lower Triangle = Women[b]							
BS		.589	.347	.060	.259	.404	.274
PS	.554		.346	.136	.239	.366	.254
VR	.440	.333		.316	.380	.271	.258
WS	.263	.069	-.020		.241	.152	.105
NGPA	.230	.169	.223	.222		.687	.328
SGPA	.282	.275	.168	.210	.711		.399
MGPA	.386	.478	.260	.175	.320	.501	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 120. [b] n = 73.

Table 56

Intercorrelations among Variables by Ethnicity for School 3

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Whites[a] ; Lower Triangle = Blacks[b]							
BS		.553	.265	.026	.028	.133	.169
PS	-.50		.266	.003	.061	.211	.237
VR	n.c.	n.c.		.111	.220	.118	.116
WS	-.50	1.00	n.c.		.194	.202	.048
NGPA	.63	-.99	n.c.	-.99		.530	.106
SGPA	.96	-.73	n.c.	-.73	.83		.216
MGPA	-.98	.67	n.c.	.67	-.78	-.99	
Upper Triangle = Asians[c] ; Lower Triangle = Hispanics[d]							
BS		.628	.513	.458	.381	.510	.221
PS	.665		.303	.405	.336	.495	.265
VR	.876	.605		.321	.438	.302	.186
WS	.431	.222	.523		.263	.016	.005
NGPA	.761	.439	.699	.560		.817	.462
SGPA	.692	.560	.407	.163	.758		.488
MGPA	.851	.879	.872	.459	.545	.505	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA. n.c. = Not computed; at least one of the variables is constant.

[a]n = 158. [b] n = 3. [c] n = 19. [d] n = 9.

Table 57

Intercorrelations among Variables for School 4

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.733	.515	.078	.245	.470	.647
PS			.532	.121	.270	.503	.541
VR				.223	.286	.319	.347
WS					.104	.051	.106
NGPA						.593	.354
SGPA							.502

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 58

Intercorrelations among Variables by Gender for School 4

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Men[a] ; Lower Triangle = Women[b]							
BS		.716	.531	.145	.211	.359	.624
PS	.688		.535	.161	.235	.445	.505
VR	.497	.551		.249	.229	.222	.332
WS	.028	.128	.193		.018	.056	.215
NGPA	.340	.394	.395	.235		.557	.239
SGPA	.557	.542	.458	.071	.676		.427
MGPA	.657	.571	.364	-.007	.497	.566	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 325. [b] n = 195.

Table 59

Intercorrelations among Variables by Ethnicity for School 4

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Whites[a] ; Lower Triangle = Blacks[b]							
BS		.649	.234	-.086	-.044	.206	.527
PS	.211		.309	.011	-.006	.316	.376
VR	.149	.084		.150	.066	.069	.032
WS	.101	.102	.213		.063	.014	.081
NGPA	-.094	.012	-.069	.021		.471	.133
SGPA	.047	-.082	-.400	-.202	.479		.461
MGPA	.396	.254	.028	.199	-.043	.068	
Upper Triangle = Asians[c] ; Lower Triangle = Hispanics[d]							
BS		.479	.256	.017	-.028	.198	.594
PS	.533		.209	.003	-.211	.191	.208
VR	.304	.381		.275	-.177	-.124	.054
WS	-.028	.104	.054		.006	-.031	.021
NGPA	-.021	.199	.142	.065		.569	-.025
SGPA	.128	.318	.137	-.016	.312		.198
MGPA	.392	.570	.332	.026	.584	.252	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 268. [b] n = 87. [c] n = 96. [d] n = 54.

Table 60

Intercorrelations among Variables for School 5

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.540	.228	.108	.023	.063	.248
PS			.254	.045	-.013	.089	.268
VR				.305	.099	-.002	.052
WS					.155	-.030	.050
NGPA						.581	.174
SGPA							303

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 61

Intercorrelations among Variables by Gender for School 5

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Men[a] ; Lower Triangle = Women[b]							
BS		.471	.258	.050	-.094	-.022	.130
PS	.588		.342	.076	-.059	.041	.198
VR	.288	.298		.259	.061	-.007	.077
WS	.225	.010	.392		.151	-.070	.061
NGPA	.243	.132	.134	.162		.513	.153
SGPA	.226	.228	-.013	.035	.680		.285
MGPA	.378	.299	.084	.039	.239	.366	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 183. [b] n = 108.

Table 62

Intercorrelations among Variables by Ethnicity for School 5

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Whites[a] ; Lower Triangle = Blacks[b]							
BS		.518	.183	.086	-.057	.008	.252
PS	-.092		.181	-.001	-.115	.043	.286
VR	-.540	.811		.235	.045	-.057	.004
WS	.000	.372	.406		.155	.001	.031
NGPA	.160	-.725	-.786	-.007		.515	.100
SGPA	.289	-.867	-.754	-.238	.607		.214
MGPA	.237	-.133	-.200	-.817	-.345	.265	
Upper Triangle = Asians[c] ; Lower Triangle = Hispanics[d]							
BS		.383	-.150	-.127	-.170	-.043	.116
PS	.795		.055	-.104	.115	.050	-.234
VR	.772	.970		.245	.091	-.246	.066
WS	-.559	-.888	-.772		-.338	-.644	-.037
NGPA	-.390	.081	-.051	-.408		.744	.277
SGPA	-.300	.226	.111	-.518	.984		.458
MGPA	-.159	.215	.324	-.066	.417	.493	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 234. [b] n = 7. [c] n = 18. [d] n = 5.

Table 63

Intercorrelations among Variables for School 6

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.543	.328	.182	-.152	.034	.392
PS			.235	.062	-.128	.097	.299
VR				.139	-.085	-.064	.157
WS					.040	-.029	.137
NGPA						.563	.115
SGPA							.298

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 64

Intercorrelations among Variables by Gender for School 6

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Men[a] ; Lower Triangle = Women[b]							
BS		.524	.304	.155	-.215	-.044	.322
PS	.529		.295	.040	-.109	.046	.274
VR	.398	.229		.189	-.119	-.119	.159
WS	.262	.183	.045		.037	-.030	.110
NGPA	-.004	-.097	-.032	.029		.602	.085
SGPA	.126	.153	.030	-.018	.516		.266
MGPA	.471	.279	.175	.207	.224	.343	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 231. [b] n = 141.

Table 65

Intercorrelations among Variables by Ethnicity for School 6

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Whites[a] ; Lower Triangle = Blacks[b]							
BS		.440	.231	.239	-.234	-.085	.317
PS	.570		.156	.022	-.241	-.056	.220
VR	.572	.246		.178	-.145	-.173	-.002
WS	.066	-.396	.164		.012	-.065	.134
NGPA	-.393	-.159	-.688	-.089		.574	.095
SGPA	-.012	-.121	-.353	-.004	.742		.241
MGPA	-.196	-.706	-.233	.609	-.086	-.135	
Upper Triangle = Asians[c] ; Lower Triangle = Hispanics[d]							
BS		.342	.153	-.015	-.283	-.028	.132
PS	.535		.149	-.078	-.315	.031	-.037
VR	.338	.196		-.113	.090	-.077	.123
WS	-.121	.148	.061		.093	-.145	.047
NGPA	-.177	-.046	.004	.047		.437	.190
SGPA	-.189	.121	-.098	.025	.571		.393
MGPA	.325	.387	.244	.004	.080	.129	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 241. [b] n = 11. [c] n = 56. [d] n = 58.

Table 66

Intercorrelations among Variables for School 7

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.662	.501	.270	.162	.395	.591
PS			.563	.321	.133	.361	.455
VR				.279	.133	.190	.333
WS					.041	.075	.223
NGPA						.568	.293
SGPA							.433

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 67

Intercorrelations among Variables by Gender for School 7

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Men[a] ; Lower Triangle = Women[b]							
BS		.609	.471	.198	.202	.340	.479
PS	.620		.605	.282	.122	.298	.280
VR	.527	.482		.308	.131	.238	.278
WS	.331	.348	.209		.048	.024	.178
NGPA	.270	.342	.187	.070		.523	.363
SGPA	.472	.467	.111	.126	.692		.408
MGPA	.638	.604	.348	.239	.306	.461	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 155. [b] n = 107.

Table 68

Intercorrelations among Variables by Ethnicity for School 7

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Whites[a] ; Lower Triangle = Blacks[b]							
BS		.481	.266	.135	-.092	.060	.350
PS	.280		.386	.182	-.174	-.061	.186
VR	.126	.090		.145	-.167	-.246	.122
WS	.095	.095	.079		-.007	-.136	.151
NGPA	-.215	-.015	.084	-.112		.423	.100
SGPA	.089	-.001	-.076	.010	.442		.253
MGPA	.566	.269	.026	-.058	.148	.027	
Upper Triangle = Asians[c] ; Lower Triangle = Hispanics[d]							
BS		.540	.421	.008	.452	.271	.504
PS	.449		.699	.158	-.027	-.091	.191
VR	.762	.513		.282	-.077	-.347	.081
WS	.345	.512	.388		-.168	-.399	-.047
NGPA	-.192	-.128	-.015	-.086		.539	.419
SGPA	-.203	.230	-.015	.111	.518		.427
MGPA	.412	.416	.617	.311	.015	.096	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 146. [b] n = 46. [c] n = 40. [d] n = 23.

Table 69

Intercorrelations among Variables for School 8

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.648	.458	.157	.231	.407	.438
PS			.478	.193	.222	.410	.403
VR				.201	.172	.289	.228
WS					.157	.145	.126
NGPA						.658	.373
SGPA							.518

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 70

Intercorrelations among Variables by Gender for School 8

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Men[a] ; Lower Triangle = Women[b]							
BS		.598	.397	.190	.282	.383	.439
PS	.680		.427	.248	.257	.415	.403
VR	.514	.565		.236	.137	.297	.221
WS	.152	.183	.177		.165	.184	.125
NGPA	.206	.248	.222	.140		.721	.318
SGPA	.420	.399	.276	.108	.599		.429
MGPA	.433	.412	.228	.137	.450	.611	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 146. [b] n = 110.

Table 71

Intercorrelations among Variables by Ethnicity for School 8

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Whites[a] ; Lower Triangle = Blacks[b]							
BS		.508	.230	.120	.046	.111	.325
PS	. .688		.286	.197	.064	.107	.286
VR	.491	.377		.208	.011	-.028	-.020
WS	.193	.102	-.112		.086	.098	.140
NGPA	.252	.229	-.016	.329		.595	.264
SGPA	.250	.312	-.117	.098	.784		.357
MGPA	.179	.196	.047	-.124	.460	.498	
Upper Triangle = Asians[c]; Lower Triangle = Hispanics[d]							
BS		.591	.115	-.278	.150	.501	.419
PS	.693		.222	-.241	.095	.250	.248
VR	-.500	.277		.023	-.196	-.195	-.233
WS	.500	.971	.500		-.132	-.168	-.237
NGPA	.751	.045	-.947	-.196		.439	.287
SGPA	-.422	-.946	-.574	-.996	.282		.583
MGPA	-.665	-.999	-.314	-.979	-.007	.958	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 181. [b] n = 27. [c] n = 33. [d] n = 3.

Table 72

Intercorrelations among Variables for School 9

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.564	.208	.064	.015	.301	.161
PS			.320	.168	.013	.338	.177
VR				.167	.145	.103	-.010
WS					.049	.039	-.083
NGPA						.428	.126
SGPA							.296

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 73

Intercorrelations among Variables by Gender for School 9

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
	Upper Triangle = Men[a] ; Lower Triangle = Women[b]						
BS		.509	.283	.076	.113	.372	.148
PS	.569		.412	.222	.083	.404	.119
VR	.153	.293		.113	.170	.136	-.038
WS	.089	.190	.226		.099	.115	-.064
NGPA	-.045	.033	.094	-.040		.410	.099
SGPA	.228	.284	.072	-.032	.480		.337
MGPA	.181	.254	.024	-.104	.164	.257	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 117. [b] n = 109.

Table 74

Intercorrelations among Variables by Ethnicity for School 9

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
			Upper Triangle = Whites[a] ; Lower Triangle = Blacks[b]				
BS		.367	-.014	-.020	-.086	.102	.015
PS	.770		.180	.096	-.067	.164	.106
VR	.362	.434		.061	.034	-.083	-.100
WS	.469	.603	.493		.020	.052	-.081
NGPA	.028	-.007	.303	.063		.501	.104
SGPA	.161	-.034	.000	-.276	.345		.289
MGPA	.010	-.149	-.463	-.223	-.093	.061	
			Upper Triangle = Asians[c] ; Lower Triangle = Hispanics[d]				
BS		.554	.246	-.086	.071	.406	.253
PS	n.c.		.030	-.121	.039	.457	.252
VR	n.c.	n.c.		.379	.272	.126	.173
WS	n.c.	n.c.	n.c.		.116	-.140	-.186
NGPA	n.c.	n.c.	n.c.	n.c.		.295	.190
SGPA	n.c.	n.c.	n.c.	n.c.	n.c.		.266
MGPA	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA. n.c. = Not computed; at least one of the variables is constant.

[a]n = 156. [b] n = 16. [c] n = 48. [d] n = 1.

Table 75

Intercorrelations among Variables for School 10

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.437	.271	.153	-.028	.083	.302
PS			.353	.173	.075	.114	.293
VR				.245	-.044	-.049	.194
WS					.071	.019	.130
NGPA						.672	.087
SGPA							.325

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 76

Intercorrelations among Variables by Gender for School 10

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Men[a] ; Lower Triangle = Women[b]							
BS		.429	.234	.105	-.044	.140	.406
PS	. .397		.330	.165	.169	.174	.349
VR	.316	.445		.185	.020	-.105	.179
WS	.197	.184	.312		.141	.083	.100
NGPA	.021	.063	-.109	.001		.684	.075
SGPA	.058	.106	-.002	-.044	.661		.383
MGPA	.221	.285	.208	.169	.095	.266	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 86. [b] n = 102.

Table 77

Intercorrelations among Variables for School 11

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.554	.247	.005	.106	.251	.386
PS			.299	.025	-.002	.207	.421
VR				.255	-.131	-.144	.207
WS					.157	.036	.112
NGPA						.594	.246
SGPA							.361

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 78

Intercorrelations among Variables by Gender for School 11

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Men[a] ; Lower Triangle = Women[b]							
BS		.486	.215	.021	.110	.176	.291
PS	.552		.291	.053	-.072	.160	.318
VR	.315	.369		.252	-.274	-.294	.147
WS	.021	.046	.258		.105	-.090	.174
NGPA	.244	.293	.069	.206		.650	.202
SGPA	.316	.244	.030	.164	.596		243
MGPA	.473	.539	.279	.063	.344	.465	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 93. [b] n = 80.

Table 79

Intercorrelations among Variables by Ethnicity for School 11

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Whites[a] ; Lower Triangle = Blacks[b]							
BS		.492	.170	-.105	-.008	-.001	.161
PS	.009		.240	-.241	-.234	.135	.460
VR	-.163	.517		.100	-.179	-.270	.175
WS	.253	.245	.372		.076	-.052	.108
NGPA	.131	-.021	.019	.376		.611	.310
SGPA	.343	-.007	-.375	.097	.579		.303
MGPA	.424	-.477	-.489	-.190	-.022	.291	
Upper Triangle = Asians[c] ; Lower Triangle = Hispanics[d]							
BS		.348	-.128	-.354	.092	.308	.478
PS	.497		.030	-.121	.039	.457	.252
VR	.457	.565		.360	-.323	-.589	-.042
WS	.243	.213	.574		.090	-.426	-.056
NGPA	-.152	.048	-.363	.279		.380	-.019
SGPA	.184	.667	.178	.373	.293		.242
MGPA	.405	.648	.600	.768	.361	.746	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 108. [b] n = 14. [c] n = 32. [d] n = 8.

Table 80

Intercorrelations among Variables for School 12

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.465	.391	.074	.102	.203	.112
PS			.259	-.062	.038	.182	.020
VR				.048	.197	.103	.203
WS					.076	.019	.086
NGPA						.617	.299
SGPA							.295

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 81

Intercorrelations among Variables by Gender for School 12

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Men[a] ; Lower Triangle = Women[b]							
BS		.433	.412	-.053	-.084	.094	.061
PS	.428		.333	-.060	-.223	-.027	-.001
VR	.393	.222		-.209	.060	.096	.157
WS	.199	-.063	.258		-.013	-.006	-.079
NGPA	.250	.232	.315	.147		.621	.380
SGPA	.222	.231	.118	.038	.623		.377
MGPA	.180	-.028	.335	.400	.173	.252	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 65. [b] n = 67.

Table 82

Intercorrelations among Variables by Ethnicity for School 12

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Whites[a] ; Lower Triangle = Blacks[b]							
BS		.356	.155	-.046	-.180	-.033	-.025
PS	.480		.109	-.162	-.167	-.006	-.132
VR	.229	.246		-.139	-.207	-.295	-.006
WS	.102	-.099	-.207		-.018	-.066	-.004
NGPA	.131	-.021	.019	.376		.511	.255
SGPA	.062	.285	.021	.131	-.044		.243
MGPA	.111	.310	.249	-.255	.026	-.327	
Upper Triangle = Asians[c] ; Lower Triangle = Hispanics[d]							
BS		.402	.550	.718	-.353	-.564	.516
PS	n.c.		-.221	.445	-.514	-.538	-.104
VR	n.c.	n.c.		.602	.195	-.187	.460
WS	n.c.	n.c.	n.c.		-.139	-.317	.421
NGPA	n.c.	n.c.	n.c.	n.c.		.678	-.143
SGPA	n.c.	n.c.	n.c.	n.c.	n.c.		.095
MGPA	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA. n.c. = Not computed; at least one of the variables is constant.

[a]n = 107. [b] n = 12. [c] n = 8. [d] n = 1.

Table 83

Intercorrelations among Variables for School 13

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.309	.152	-.013	-.104	.000	.207
PS			.226	-.053	-.247	.121	.070
VR				.180	-.085	-.027	.092
WS					.058	-.050	.056
NGPA						.527	.037
SGPA							.228

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 84

Intercorrelations among Variables by Gender for School 13

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Men[a] ; Lower Triangle = Women[b]							
BS		.252	.065	-.007	.019	.059	.049
PS	.312		.159	.001	-.263	.075	-.110
VR	.221	.299		.129	.059	.003	-.035
WS	.041	-.087	.309		.077	-.059	.151
NGPA	-.164	-.023	-.293	-.103		.508	.093
SGPA	.023	.374	-.027	-.108	.491		.304
MGPA	.485	.479	.323	-.169	-.124	.076	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 92. [b] n = 48.

Table 85

Intercorrelations among Variables by Ethnicity for School 13

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Whites[a] ; Lower Triangle = Blacks[b]							
BS		.292	.045	-.094	-.078	-.025	.241
PS	-.033		.200	.009	-.240	.072	.037
VR	.537	.083		.141	-.100	-.064	.112
WS	.737	-.224	.486		.059	-.040	.102
NGPA	-.598	.187	-.388	-.900		.541	.117
SGPA	-.097	.664	-.278	-.339	.519		.234
MGPA	-.072	.680	-.478	-.289	.295	.883	
Upper Triangle = Asians[c] ; Lower Triangle = Hispanics[d]							
BS		.795	.652	.057	-.665	-.429	.358
PS	-.997		.543	-.382	-.484	-.050	.039
VR	.127	-.052		.238	-.261	-.166	.004
WS	.524	-.459	.911		-.284	-.637	-.044
NGPA	.306	-.234	.983	.971		.686	-.337
SGPA	.063	.012	.998	.883	.969		-.150
MGPA	.925	-.951	-.260	.161	-.079	-.322	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 117. [b] n = 7. [c] n = 10. [d] n = 3.

Table 86

Intercorrelations among Variables for School 14

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
BS		.444	.249	.088	-.066	-.128	.323
PS			.263	.113	-.020	.092	.364
VR				.317	-.162	-.288	.063
WS					.065	-.237	.094
NGPA						.497	.039
SGPA							.184

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

Table 87

Intercorrelations among Variables by Gender for School 14

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
Upper Triangle = Men[a] ; Lower Triangle = Women[b]							
BS		.473	.303	.020	-.194	-.101	.547
PS	.416		.252	.079	-.063	-.001	.466
VR	.243	.343		.247	-.151	-.396	.311
WS	.169	.207	.354		.368	-.076	.113
NGPA	.052	.064	-.222	-.187		.356	-.047
SGPA	-.151	.185	-.208	-.333	.605		.096
MGPA	.172	.286	-.194	.069	.094	.244	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 41. [b] n = 48.

Table 88

Intercorrelations among Variables by Ethnicity for School 14

Variables	BS	PS	VR	WS	NGPA	SGPA	MGPA
			Upper Triangle = Whites[a] ; Lower Triangle = Blacks[b]				
BS		.282	-.033	.168	.986	-.177	.766
PS	.327		.530	.649	.198	.382	.264
VR	.217	.196		.760	-.053	.693	.105
WS	.024	.061	.309		.160	.116	-.084
NGPA	-.266	-.105	-.188	.014		-.242	.683
SGPA	-.235	.014	-.407	-.254	.568		.352
MGPA	.152	.209	-.038	.093	-.029	.116	
			Upper Triangle = Asians[c] ; Lower Triangle = Hispanics[d]				
BS		.519	-.051	.293	.347	-.125	.332
PS	.971		.088	.476	.246	-.167	.775
VR	.866	.961		.083	-.215	-.742	-.118
WS	-.945	-.839	-.655		.360	-.359	.297
NGPA	-.459	-.659	-.842	.143		.284	-.057
SGPA	.852	.701	.476	-.976	.075		-.097
MGPA	1.000	.965	.854	-.953	-.437	.864	

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; NGPA =Undergraduate non-science GPA; SGPA =Undergraduate science GPA; MGPA = First-year medical school GPA.

[a]n = 6. [b] n = 68. [c] n = 11. [d] n = 3.

Appendix E

Standard Errors of Estimate, Regression Slopes, and Regression Intercepts by Gender and Ethnicity

Table 89

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 1

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	.60	.08	2.04	.69	.08	2.04
PS	.63	.09	1.95	.70	.06	2.23
VR	.63	.09	1.89	.70	.07	2.09
WS	.64	.06	2.43	.70	-.02	2.84

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 90

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Ethnicity for School 1

	Whites			Blacks		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	.63	.09	1.92	.80	.10	1.91
PS	.62	.11	1.80	.81	.01	2.57
VR	.63	.11	1.72	.81	-.02	2.77
WS	.64	.05	2.50	.76	-.16	3.52

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 91

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 2

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	.86	.30	4.80	.85	.48	2.71
PS	.84	.28	5.07	.91	.38	3.93
VR	.88	.30	.52	.89	.40	3.85
WS	1.01	.06	7.70	1.16	.11	7.03

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 92

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Ethnicity for School 2

	Whites			Asians		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	.59	.18	6.40	.72	.24	5.35
PS	.61	.13	7.01	.74	.18	1.21
VR	.64	.30	7.76	.78	.09	7.18
WS	.64	.05	8.12	.67	-.26	9.88

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 93

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 3

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	5.98	1.07	74.46	5.49	1.32	71.97
PS	6.01	.93	75.80	5.23	1.95	67.78
VR	6.01	.87	76.17	5.75	.71	76.84
WS	6.06	.39	81.17	5.86	.61	79.12

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 94

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 4

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	7.08	2.68	45.31	8.05	3.40	39.60
PS	7.82	1.95	51.78	8.77	2.99	43.55
VR	8.55	1.54	56.51	9.95	1.98	49.76
WS	8.85	.97	64.82	10.68	-.04	67.23

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 95

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Ethnicity for School 4

	Whites			Blacks			Asians			Hispanics		
Variable	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0
BS	7.23	2.46	48.91	7.62	2.28	45.48	6.52	2.83	43.26	9.07	2.29	46.80
PS	7.89	1.50	57.94	8.02	2.10	46.77	7.93	.94	62.45	8.09	2.65	43.99
VR	8.51	.17	70.69	8.29	.14	59.10	8.09	.29	69.40	9.29	1.66	51.21
WS	8.49	.34	70.34	8.13	.80	55.60	8.10	.08	71.62	9.85	.12	62.97

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 96

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 5

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	.56	.05	2.64	.51	.14	1.75
PS	.55	.06	2.52	.53	.10	2.14
VR	.56	.03	2.87	.55	.03	2.65
WS	.56	.02	3.02	.55	.01	2.87

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 97

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 6

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	.70	.17	1.54	.64	.21	1.03
PS	.72	.12	1.98	.70	.14	1.81
VR	.74	.08	2.36	.71	.08	2.16
WS	.75	.04	2.88	.71	.08	2.43

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 98

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Ethnicity for School 6

	Whites			Asians			Hispanics		
Variable	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0
BS	.65	.16	1.71	.67	.08	2.28	.73	.16	1.29
PS	.67	.10	2.33	.67	-.02	3.18	.71	.18	1.23
VR	.69	-.01	3.22	.67	.06	2.49	.74	.12	1.55
WS	.68	.05	2.92	.67	.02	2.93	.77	.02	2.61

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 99

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 7

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	.69	.21	.55	.56	.26	-.02
PS	.75	.11	1.61	.58	.26	.09
VR	.75	.12	1.52	.68	.14	.95
WS	.77	.07	2.15	.71	.10	1.56

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 100

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Ethnicity for School 7

	Whites			Blacks			Asians		
Variable	SEE	B_1	B_0	SEE	B_1	B_0	SEE	B_1	B_0
BS	.64	.16	1.01	.64	.27	-.13	.57	.24	.46
PS	.67	.07	1.89	.75	.19	.50	.65	.07	2.12
VR	.67	.06	2.03	.77	.01	1.65	.66	.03	2.56
WS	.67	.06	2.21	.77	-.03	2.97	.66	-.02	2.97

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 101

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 8

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	1.61	.46	4.40	1.76	.46	4.55
PS	1.64	.36	5.32	1.78	.45	4.70
VR	1.75	.28	6.47	1.90	.24	6.68
WS	1.78	.11	8.57	1.93	.13	8.14

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 102

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Ethnicity for School 8

	Whites			Asians		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	1.59	.34	5.87	1.47	.442	4.52
PS	1.61	.27	6.55	1.57	.25	6.54
VR	1.68	-.03	9.78	1.58	-.31	12.28
WS	1.67	.11	8.68	1.58	-.19	10.34

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 103

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 9

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	.69	.06	1.86	.70	.08	1.76
PS	.70	.05	2.03	.69	.10	1.50
VR	.70	-.02	2.81	.71	.01	2.47
WS	.70	-.03	2.79	.71	-.04	2.92

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 104

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Ethnicity for School 9

	Whites			Asians		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	.68	.01	2.59	.69	.11	1.36
PS	.67	.05	2.13	.69	.12	1.13
VR	.68	-.05	3.23	.70	.13	1.28
WS	.68	-.03	2.90	.70	-.08	3.158

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 105

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 10

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	5.50	1.67	67.06	4.98	.75	76.69
PS	5.64	1.13	72.48	4.90	1.02	75.06
VR	5.93	.68	76.81	4.99	.72	76.69
WS	5.49	.35	81.01	5.04	.46	80.52

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 106

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 11

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	.68	.14	1.03	.70	.24	.07
PS	.67	.12	1.09	.67	.24	.05
VR	.70	.06	1.84	.76	.13	1.03
WS	.70	.07	1.91	.79	.02	2.11

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 107

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Ethnicity for School 11

	Whites			Asians		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	.71	.08	1.60	.56	.20	.49
PS	.69	.11	1.30	.56	.15	.89
VR	.71	.08	1.67	.63	-.01	2.60
WS	.71	.04	2.11	.63	-.02	2.59

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 108

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 12

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	.61	.02	2.76	.30	.04	2.57
PS	.61	-.01	2.98	.30	-.06	2.94
VR	.60	.06	2.43	.27	.06	2.32
WS	.61	-.03	3.14	.28	.06	2.52

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 109

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 13

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	.66	.03	2.70	.58	.19	1.41
PS	.66	.05	3.32	.58	.21	1.36
VR	.66	-.02	3.07	.63	.12	1.85
WS	.65	.05	2.66	.65	-.06	3.33

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 110

Standard Errors of Estimate, Regression Slopes, and Intercepts of MCAT Subtest Scores by Gender for School 14

	Men			Women		
Variable	SEE	B_1	B_0	SEE	B_1	B_0
BS	6.35	2.82	56.76	7.81	.86	72.10
PS	6.71	2.09	62.01	7.59	1.91	64.79
VR	7.21	1.02	69.94	7.78	-.83	84.66
WS	7.54	.48	75.07	7.91	.26	76.68

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 111

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores for School 1

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	.62		.08		.06	1.63
Women	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
Whites	.61		.09	.08		1.11
Blacks	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept

n.c. = Not computed; none of the predictors were entered into the regression model.

Table 112

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores for School 2

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	.80	.20		.18		4.21
Women	.80	.33		.22		2.26
Whites	.60	.18				6.40
Asians	.62	.22			-.24	7.31

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 113

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores for School 3

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	5.98	1.07				74.47
Women	5.23		1.95			67.78

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 114

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores for School 4

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	7.08	2.69				45.31
Women	7.73	2.53	1.47			34.60
Whites	7.23	2.46				48.91
Blacks	7.28	2.51	2.57			27.81
Asians	6.52	2.83				43.26
Hispanics	8.09		2.65			43.99

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 115

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores for School 5

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	.55		.07			2.52
Women	.51	.14				1.75

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 116

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores for School 6

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	.71	.13	.06			1.54
Women	.64	.21				1.03
Whites	.65	.16				1.71
Asians	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
Hispanics	.71		.18			1.23

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept

n.c. = Not computed; none of the predictors were entered into the regression model.

Table 117

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores for School 7

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	.69	.21				.55
Women	.53	.18	.14			.46
Whites	.64	.16				1.01
Blacks	.64	.27				.13
Asians	.57	.24				.46

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 118

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores by Gender for School 8

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	1.58	.32	.20			3.67
Women	1.76	.46				4.55
Whites	1.58	.26	.16			5.11
Asians	1.47	.44				4.52

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 119

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores by Gender for School 9

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
Women	.69	.10				1.50
Whites	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
Asians	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept

n.c. = Not computed; none of the predictors were entered into the regression model.

Table 120

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores for School 10

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	5.50	1.67				67.06
Women	4.90		1.02			75.06

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 121

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores for School 11

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	.67		.12			1.09
Women	.65	.13	.18			-.57
Whites	.69		.11			1.30
Asians	.53	.16	.11			-.25

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Table 122

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores for School 12

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
Women	.27			.05	.05	2.15

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept

n.c. = Not computed; none of the predictors were entered into the regression model.

Table 123

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores for School 13

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
Women	.54	.14	.16			.54

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept

n.c. = Not computed; none of the predictors were entered into the regression model.

Table 124

Standard Errors of Estimate, Regression Slopes, and Intercepts for the Combination of MCAT Subtest Scores for School 14

		B_1				
Group	SEE	BS	PS	VR	WS	B_0
Men	6.35		2.81			56.76
Women	7.26		2.67	-1.41		70.66

Note. BS = Biological Sciences; PS = Physical Sciences; VR = Verbal Reasoning; WS = Writing Sample; SEE = Standard error of estimate; B_1 = Regression Slope; B_0 = Regression Intercept.

Appendix F

Mean Residual Values for Men and Women and for White, Black, Asian, and Hispanic Students Using a Single Regression Equation for All Students

Table 125

Mean Residual Values for Men and Women Using the Total-Group Regression Equation

	Men					Women				
School	BS	PS	VR	WS	MCATs	BS	PS	VR	WS	MCATs
1	.03	.01	.02	.03	-.01	-.01	-.02	-.03	-.05	.02
2	.04	.03	.18	.17	.05	-.03	-.03	-.22	-.19	-.07
3	.12	-.09	.34	.40	-.07	-.20	.14	-.57	-.66	.11
4	-.06	-.15	1.06	1.27	-.27	.11	.25	-1.85	-2.20	.47
5	.04	.03	.07	.07	.03	-.08	-.05	-.12	-.11	-.04
6	.04	.02	.08	.07	.01	-.05	-.02	-.12	-.13	-.03
7	.58	.08	.16	.49	.05	.39	-.13	-.23	.04	-.06
8	-.04	-.11	.07	.11	-.10	.07	.15	-.09	-.15	.15
9	.04	-.03	-.05	-.02	-.03	.08	.04	-.01	.01	.04
10	-.47	-.75	-.14	-.81	-.77	.40	.64	.12	-.42	.66
11	-.02	-.06	.05	.06	-.07	.03	.06	-.06	-.05	.08
12	.03	.04	.05	.04	.05	-.04	-.04	-.05	-.05	-.05
13	-.04	-.03	-.02	-.01	-.04	.07	.04	.04	.02	.07
14	-.71	-.82	.34	-.28	-.82	.59	.68	.28	.23	.68
Weighted	.03	-.08	.17	.18	-.09	.09	.11	-.22	-.28	.13

Table 126

Mean Residual Values for Whites and Blacks Using the Total-Group Regression Equation

	Whites					Blacks				
School	BS	PS	VR	WS	MCATs	BS	PS	VR	WS	MCATs
1	.03	.39	.03	.05	.03	.03	.03	.03	-.12	.03
4	1.25	1.69	1.84	3.23	1.11	-1.85	-3.31	-5.58	-8.93	-1.30
7	.56	.05	.06	.44	.03	.20	-.30	-.44	-.35	-.16
Weighted	.58	.66	.58	1.13	.35	-.57	-1.29	-2.16	-3.34	-.52

Table 127

Mean Residual Values for Whites and Asians Using the Total-Group Regression Equation

	Whites					Asians				
School	BS	PS	VR	WS	MCATs	BS	PS	VR	WS	MCATs
2	.27	.34	.24	.51	.19	-.05	-.19	.03	.13	-.15
4	1.25	1.69	1.84	3.23	1.11	-.84	-.08	2.29	2.98	-1.14
6	.11	.13	.14	.15	.10	-.08	-.10	-.01	-.03	-.11
7	.56	.05	.06	.44	.03	.73	.22	.39	.71	.17
8	.20	.17	.18	.27	.17	-.14	-.06	.10	-.02	-.13
9	.11	.05	.06	.06	.05	.03	-.03	-.01	-.01	-.03
11	.05	.06	.05	.06	.05	.09	.02	.18	.17	.03
Weighted	.34	.31	.33	.60	.22	-.04	-.03	.04	.56	-.20

Table 128

Mean Residual Values for Whites and Hispanics Using the Total-Group Regression Equation

	Whites					Hispanics				
School	BS	PS	VR	WS	MCATs	BS	PS	VR	WS	MCATs
4	1.25	1.69	1.84	3.23	1.11	-1.09	-2.12	-3.12	-5.34	-.82
6	.11	.13	.14	.15	.10	-.26	-.30	-.39	-.43	-.24
Weighted	.50	.67	.73	1.21	.45	-.54	-.91	-1.31	-2.09	-.44

Appendix G

Number and Percentages of Students Whose First-Year Medical School GPAs were Over/Underpredicted by MCAT Subtest Scores, Individually and in Combination

Table 129

Number and Percentages of Men and Women Whose First-Year Medical School GPAs were Over/Underpredicted by Biological Sciences Subtest Scores

	Men				*Women*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
1	58	79	42.3	57.7	28	31	47.5	52.5
2	26	34	43.3	56.7	25	26	46.0	51.0
3	56	63	47.1	52.9	33	38	46.5	53.5
4	71	80	47.0	53.0	38	49	43.7	56.3
5	78	98	44.3	55.7	66	41	61.7	38.3
6	92	136	40.4	59.6	68	69	49.6	50.4
7	28	124	18.4	81.6	24	80	23.1	76.9
8	60	80	44.4	55.6	53	56	48.6	51.4
9	56	59	48.7	51.3	53	56	48.6	51.4
10	41	42	49.4	50.6	39	58	40.2	59.8
11	51	35	59.3	40.7	44	33	57.1	42.9
12	19	46	29.2	70.8	18	48	27.3	72.7
13	37	53	41.1	58.9	16	28	36.4	63.6
14	19	21	47.5	52.5	17	31	35.4	64.6
Mean	49	67	43.0	57.0	37	46	43.7	56.3

Table 130

Number and Percentages of Men and Women Whose First-Year Medical School GPAs were Over/Underpredicted by Physical Sciences Subtest Scores

	Men				*Women*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
1	83	54	60.6	39.4	27	32	45.8	54.2
2	25	35	41.7	58.3	26	25	51.0	49.0
3	63	56	58.3	41.7	31	40	43.7	56.3
4	71	80	47	53	36	51	41.4	58.6
5	79	97	55.1	44.9	58	49	54.2	45.8
6	138	90	60.5	39.5	67	70	48.6	51.1
7	64	88	42.1	57.9	57	47	54.8	45.2
8	74	70	48.6	51.4	49	60	45.0	55.0
9	59	56	51.3	48.7	52	57	47.7	52.3
10	42	41	50.6	49.4	38	59	39.2	60.8
11	49	37	57	43	41	36	53.2	46.8
12	13	52	20	80	16	50	24.2	75.8
13	52	38	57.8	42.2	18	26	40.9	59.1
14	21	19	52.5	47.5	13	35	27.1	72.9
Mean	60	58	50.3	49.7	38	46	44.1	55.9

Table 131

Number and Percentages of Men and Women Whose First-Year Medical School GPAs were Over/Underpredicted by Verbal Reasoning Subtest Scores

	Men				*Women*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
1	57	80	41.6	58.4	28	31	45.8	54.2
2	25	35	41.7	58.3	28	23	54.9	45.1
3	49	70	40.8	58.3	34	37	47.9	52.1
4	63	88	41.7	58.3	46	41	52.9	47.1
5	76	100	43.2	56.8	63	44	58.9	41.1
6	84	144	36.8	63.2	69	68	50.4	49.6
7	62	90	40.8	59.2	67	37	64.4	35.6
8	65	79	45.1	54.9	64	45	58.7	41.3
9	59	56	51.3	48.7	55	54	50.5	49.5
10	40	43	48.2	51.8	40	57	41.2	58.8
11	45	41	52.3	47.7	46	31	59.7	40.3
12	28	37	43.1	56.9	32	34	48.5	51.5
13	37	53	41.1	58.9	16	28	36.4	63.6
14	16	24	40.0	60.0	17	31	35.4	64.6
Mean	50	67	43.4	56.6	43	40	50.4	49.6

Table 132

Number and Percentages of Men and Women Whose First-Year Medical School GPAs were Over/Underpredicted by Writing Sample Subtest Scores

	Men				*Women*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
1	58	79	42.3	57.7	31	28	47.5	52.5
2	20	40	33.3	66.7	28	23	54.9	45.1
3	50	69	42.0	58.0	36	36	50.0	50.0
4	63	88	41.7	58.3	51	36	58.6	41.4
5	75	101	42.6	57.4	65	42	60.7	39.3
6	88	140	38.6	61.4	73	64	53.3	46.7
7	41	111	27.0	73.0	50	54	48.1	51.9
8	64	80	44.4	55.6	59	50	54.1	45.9
9	55	60	47.8	52.2	55	54	50.5	49.5
10	37	46	43	53.5	42	55	43.3	56.7
11	45	41	52.3	47.7	44	33	57.1	42.9
12	14	51	21.5	78.5	18	48	27.3	72.7
13	39	51	43.3	56.7	17	27	38.6	61.4
14	16	24	40.0	60.0	18	30	37.5	62.5
Mean	48	70	40.2	59.8	42	41	48.7	51.3

Table 133

Number and Percentages of Men and Women Whose First-Year Medical School GPAs were Over/Underpredicted by Combination of MCAT Subtest Scores

	Men				*Women*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
1	75	62	54.7	45.3	26	33	52.5	47.5
2	26	34	43.3	56.7	29	22	56.9	43.1
3	56	63	47.1	52.9	30	41	42.3	57.7
4	73	78	48.3	51.7	39	48	44.8	55.2
5	77	99	43.8	56.2	57	50	53.3	46.7
6	137	91	60.1	39.9	64	73	46.7	53.3
7	66	86	43.4	56.6	54	50	51.9	48.1
8	76	68	52.8	47.2	49	60	45.0	55.0
9	59	56	51.3	48.7	52	57	47.7	52.3
10	43	40	51.8	48.2	37	60	38.1	61.9
11	51	35	59.3	40.7	42	35	54.5	45.5
12	65	0	100	0	0	67	0.0	100.0
13	53	37	58.9	41.1	16	28	36.4	63.6
14	21	19	52.5	47.5	13	35	27.1	72.9
Mean	63	55	54.8	45.2	36	47	42.7	57.3

Table 134

Number and Percentages of White and Black Students Whose First-Year Medical School GPAs were Over/Underpredicted by Biological Sciences Subtest Scores

	White				*Black*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
1	62	87	41.6	58.4	15	20	60	40
4	49	76	39.2	60.8	21	17	55.3	44.7
7	26	119	17.9	82.1	15	29	34.1	65.9
Mean	46	94	32.9	67.1	17	22	49.8	50.2

Table 135

Number and Percentages of White and Black Students Whose First-Year Medical School GPAs were Over/Underpredicted by Physical Sciences Subtest Scores

	White				*Black*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
1	57	92	38.3	61.7	10	15	40.0	60.0
4	48	77	38.4	61.6	22	16	57.9	42.1
7	65	80	44.5	55.2	28	16	63.6	36.4
Mean	57	83	40.4	59.5	20	16	53.8	46.2

Table 136

Number and Percentages of White and Black Students Whose First-Year Medical School GPAs were Over/Underpredicted by Verbal Reasoning Subtest Scores

	White				*Black*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
1	62	87	41.6	58.4	10	15	40.0	60.0
4	47	78	37.6	62.4	26	12	68.4	31.6
7	70	75	48.3	51.7	31	13	70.5	29.5
Mean	60	80	42.5	57.5	22	13	56.6	40.4

Table 137

Number and Percentages of White and Black Students Whose First-Year Medical School GPAs were Over/Underpredicted by Writing Sample Subtest Scores

	White				*Black*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
1	63	86	42.3	57.7	12	13	48.0	52.0
4	43	82	34.4	65.6	35	5	92.1	7.9
7	44	101	30.3	69.7	30	14	68.2	31.8
Mean	50	90	35.7	64.3	26	11	42.7	57.3

Table 138

Number and Percentages of White and Black Students Whose First-Year Medical School GPAs were Over/Underpredicted by Combination of MCAT Subtest Scores

	White				*Black*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
1	66	83	44.3	55.7	16	9	64	36
4	52	73	41.6	58.4	20	18	52.6	47.4
7	69	76	47.6	52.4	25	19	56.8	43.2
Mean	62	77	45.5	55.5	20	15	57.8	42.2

Table 139

Number and Percentages of White and Asian Students Whose First-Year Medical School GPAs were Over/Underpredicted by Biological Sciences Subtest Scores

	White				*Asian*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
2	17	31	35.4	64.6	13	15	46.4	53.6
4	49	76	39.2	60.8	17	23	42.5	57.5
6	89	149	37.4	62.6	26	29	47.3	52.7
7	26	119	17.9	82.1	3	36	7.7	92.3
8	71	109	39.4	60.6	19	14	57.6	42.4
9	72	83	46.5	53.5	24	23	51.1	48.9
11	63	41	60.6	39.4	12	19	38.7	61.3
Mean	55	87	39.5	60.5	16	23	41.6	58.4

Table 140

Number and Percentages of White and Asian Students Whose First-Year Medical School GPAs were Over/Underpredicted by Physical Sciences Subtest Scores

	White				*Asian*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
2	13	35	27.1	72.9	18	10	64.3	35.7
4	48	77	38.4	61.6	18	22	45.0	55.0
6	80	158	33.6	66.4	30	25	54.6	45.5
7	65	80	44.5	55.2	14	25	35.9	64.1
8	78	102	43.3	56.7	17	16	51.5	48.5
9	74	81	47.4	51.9	24	23	51.1	48.9
11	56	48	51.9	44.4	15	16	48.4	51.6
Mean	59	83	40.9	58.1	20	19	50.1	49.9

Table 141

Number and Percentages of White and Asian Students Whose First-Year Medical School GPAs were Over/Underpredicted by Verbal Reasoning Subtest Scores

	White				*Asian*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
2	17	31	35.4	64.6	13	15	46.4	53.6
4	47	78	37.6	62.4	14	26	35.0	65.0
6	79	159	33.2	66.8	25	30	45.5	54.5
7	70	75	48.3	51.7	11	28	27.5	70.0
8	83	97	46.1	53.9	17	16	51.5	48.5
9	76	79	49.0	51.0	25	22	53.2	46.8
11	57	47	54.8	45.2	13	18	41.9	58.1
Mean	61	81	43.5	56.5	17	22	43.4	56.6

Table 142

Number and Percentages of White and Asian Students Whose First-Year Medical School GPAs were Over/Underpredicted by Writing Sample Subtest Scores

	White				*Asian*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
2	10	38	20.8	79.2	12	16	42.9	57.1
4	43	82	34.4	65.6	11	29	27.5	72.5
6	83	155	34.9	65.1	26	29	47.3	52.7
7	44	101	30.3	69.7	6	33	15.4	84.6
8	78	102	43.3	56.7	16	17	48.5	51.5
9	73	82	47.1	52.9	24	23	51.1	48.9
11	56	48	51.9	44.4	12	19	38.7	61.3
Mean	55	87	38.7	61.3	15	23	38.8	61.2

Table 143

Number and Percentages of White and Asian Students Whose First-Year Medical School GPAs were Over/Underpredicted by Combination of MCAT Subtest Scores

	White				*Asian*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
2	18	30	37.5	62.5	18	10	64.3	35.7
4	52	73	41.6	58.4	20	20	50.0	50.0
6	86	152	36.1	63.9	27	28	49.1	50.9
7	69	76	47.6	52.4	13	26	33.3	66.7
8	74	106	41.1	58.9	18	15	54.5	45.5
9	74	81	47.4	51.9	24	23	51.1	48.9
11	60	44	57.7	42.3	15	16	48.4	51.6
Mean	62	80	44.2	55.8	20	19	50.1	49.9

Table 144

Number and Percentages of White and Hispanic Students Whose First-Year Medical School GPAs were Over/Underpredicted by Biological Sciences Subtest Scores

	White				*Hispanic*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
4	49	76	39.2	60.8	18	11	62.1	37.9
6	89	149	37.4	62.6	34	23	59.6	40.4
Mean	69	113	38.3	61.7	26	17	60.9	39.1

Table 145

Number and Percentages of White and Hispanic Students Whose First-Year Medical School GPAs were Over/Underpredicted by Physical Sciences Subtest Scores

	White				*Hispanic*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
4	48	77	38.4	61.6	16	13	55.2	44.8
6	80	158	33.6	66.4	35	22	61.4	38.6
Mean	64	118	36.0	64.0	26	18	58.3	41.7

Table 146

Number and Percentages of White and Hispanic Students Whose First-Year Medical School GPAs were Over/Underpredicted by Verbal Reasoning Subtest Scores

	White				Hispanic			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
4	47	78	37.6	62.4	18	11	62.1	37.9
6	79	159	33.2	66.8	38	19	66.7	33.3
Mean	63	119	35.4	64.6	28	15	64.4	35.6

Table 147

Number and Percentages of White and Hispanic Students Whose First-Year Medical School GPAs were Over/Underpredicted by Writing Sample Subtest Scores

	White				*Hispanic*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
4	43	82	34.4	65.6	21	8	72.4	27.6
6	83	155	34.9	65.1	39	18	68.4	31.6
Mean	63	119	34.6	65.3	30	13	70.4	29.6

Table 148

Number and Percentages of White and Hispanic Students Whose First-Year Medical School GPAs were Over/Underpredicted by Combination of MCAT Subtest Scores

	White				*Hispanic*			
School	*Over*	*Under*	*% Over*	*% Under*	*Over*	*Under*	*% Over*	*% Under*
4	52	73	41.6	58.4	16	13	55.2	44.8
6	86	152	36.1	63.9	32	25	56.1	43.9
Mean	69	113	38.8	61.2	24	19	55.6	44.4

Appendix H

MCAT Writing Sample Rubric

Table 149

Description of Writing Skills Associated with Writing Sample Alphabetic Scores

J	K	L	M	N	O	P	Q	R	S	T
Above Average			Average					Above Average		
These essays demonstrate a degree of difficulty in discussing the topic and/or responding to the three writing tasks. They may show problems with the mechanics of writing or in addressing the topic at a complex level. The response may not be integrated. Although the three writing tasks may be addressed, the ideas may be underdeveloped.			These essays demonstrate a degree of proficiency in discussing the topic and/or responding to the three writing tasks. Few problems in the mechanics of writing are evident, and there is demonstration of control of language. The writing tasks are addressed in clear, organized, and coherent manner. Ideas are developed to some extent and may show evidence of depth and complexity of thought.					These essays respond to the topic and the three writing tasks in a superior manner. The writing demonstrates strong control of language. The response is presented in a clear, organized, and coherent fashion. Ideas are well-developed, and the topic is dealt with at a complex level.		

Adapted from "Scoring the MCAT Writing Sample: Examples of MCAT Writing Sample Responses and Explanations of their Scores," by Association of American Colleges, 2003, http://www.aamc.org/students/mcat/studentmanual/writingsample/scoring.pdf. Copyright 2003 by Association of American Colleges. Medical College Admission Test.

Appendix I

Formulae for Correcting for Restriction of Range and Criterion Unreliability

Correction for Restriction of Range

Thorndike (1949) described three situations for correcting for restriction of range.

Case 1: The values of the standard deviation for variable 2 are known for both the unrestricted and restricted groups.

Thorndike described this case as "one that is not likely to be encountered often in practice" (p. 173). He illustrated this case as follows: A research test (variable 2) is given to a random sample of applicants and the standard deviation of the research test for this sample is determined. Later, the research test is correlated with a selection test (variable 1). The correction formula for restriction of range in such a situation is given as:

$$R_{12} = \sqrt{1 - \frac{s_2^{\,2}}{S_2^{\,2}}\left(1 - r_{12}^{\,2}\right)}$$

R_{12} and r_{12} are the correlations between variable 1 and 2 in the unrestricted and restricted groups, respectively, and S_2 and s_2 are the standard deviations of variable 2 in the unrestricted and restricted groups, respectively.

Case 2: The values of the standard deviation for variable 1 are known for both the unrestricted and restricted groups.

Thorndike explained that:

> This case has a fair amount of practical significance because it is encountered whenever we wish to obtain an estimate of the validity of a selection procedure which we have actually been using as applied to the general group of applicants for a job category or school (p. 174).

In Case 2, a selection test (variable 1) is administered to applicants, who are later selected based on their test score. The test scores for the selected applicants are then correlated with applicants' scores on a criterion measure (variable 2). In order to obtain the "actual" correlation between the selection test and the criterion measure the following correction formula is applied:

$$R_{12} = \frac{r_{12}\dfrac{S_1}{s_1}}{\sqrt{1 - r_{12}^2 + r_{12}^2\dfrac{S_1^2}{s_1^2}}}$$

Again, R_{12} and r_{12} are the correlations between variables 1 and 2 in the unrestricted and restricted groups, respectively, and S_1 and s_1 are the standard deviations of variable 1 in the unrestricted and restricted groups, respectively.

Case 3: *The values of the standard deviation for variable 3 are known for both the unrestricted and restricted groups.*

In this third case, a third variable (variable 3) is introduced. This case occurs when a test (variable 3) is administered to applicants and their scores are recorded. A fraction of the applicants are then selected based on their scores on variable 3. Later, the selected applicants' scores on another selection test (variable 1) and a criterion measure (variable 2) are correlated. To compensate for restriction of range on this correlation coefficient the following formula was proposed:

$$R_{12} = \frac{r_{12} + r_{13}r_{23}\left(\dfrac{S_3^2}{s_3^2} - 1\right)}{\sqrt{\left[1 + r_{13}^2\left(\dfrac{S_3^2}{s_3^2} - 1\right)\right]\left[1 + r_{23}^2\left(\dfrac{S_3^2}{s_3^2} - 1\right)\right]}}$$

R_{12}, r_{12}, r_{13}, and r_{23} are the correlations between variables 1 and 2 in the unrestricted group and the correlations between variables 1 and 2, variables 1 and 3, and variables 2 and 3 in the restricted group, respectively. S_3 and s_3 are the standard deviations of variable 3 in the unrestricted and restricted groups, respectively.

Correction for Criterion Unreliability

Thorndike (1949) mathematically defined the relationship of the correlation between two hypothetical perfectly reliable measures to the correlation obtained from two sets of actual observations as:

$$r_{A\infty B\infty} \frac{r_{AB}}{\sqrt{r_{AA} r_{BB}}}$$

Thorndike rewrote this formula as:

$$r_{AB} = r_{A\infty B\infty} \sqrt{r_{AA} r_{BB}}$$

r_{AB} is the correlation between the actual scores on A and B, $r_{A\infty B\infty}$ is the correlation between perfectly reliable "true" scores on variable A and B, and r_{AA} and r_{BB} are the reliabilities of the measures of variables A and B, respectively. According to Thorndike, "this is the formula to estimate true score correlation from the correlation of the fallible measures or to correct correlation coefficients for attenuation due to unreliability of measurement" (p. 105).

Similarly, Lord and Novick's (1968) equation:

$$\rho\left(T_x, T_y\right) = \frac{\rho_{xy}}{\sqrt{\rho_{xx'}\, \rho_{yy'}}}$$

where ρ (T_X, T_Y) represents the correlation between the true scores of variables X and Y, ρ_{xy} represents the correlation between the actual scores on variables X and Y, and $\rho_{XX'}$ and $\rho_{YY'}$ represent the reliabilities of measures of X and Y, respectively, explains the correlation between true scores in terms of the correlation between observed scores and the reliability of each measurement.

In event where only the criterion is corrected for unreliability Thorndike (1949) proposed the following formula:

$$r_{A\infty B} = \frac{r_{AB}}{\sqrt{r_{AA}}}$$

to correct observed validity coefficients for criterion unreliability. According to Thorndike, this formula estimates “what the validity would have been if the criterion had been perfectly reliable” (p. 107).

Appendix J

Procedure for Performing the Chow Test in SPSS

For a given dependent variable, Y, a grouping variable, Group, and a predictor X, the procedure for conducting the Chow test in SPSS are outlined as follows (Matheson, 2001):

First, the SPSS pull-down menus, Analyze, General Linear Model, and Univariate, are selected simultaneously. Then, the dependent variable, Y, the grouping variable, Group, and the predictor variable, X, are moved into the "Dependent Variable", "Fixed Factor(s)", and Covariate(s) boxes, respectively. Second, the model is specified as custom model which includes the Group*X or interaction between the grouping variable and the predictor. The Group and Group*X terms test for whether the intercepts and slopes, respectively, differ among the different groups.