THE EFFECTS OF TEST PREPARATION METHODS EMPLOYED BY MEDICAL STUDENTS ON STANDARDIZED ADMISSION TESTS

by

EDNA E. JACKSON-GRAY

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2000

MAJOR: EVALUATION AND RESEARCH

Approved by:

Advisor Date

DEDICATION

This research is dedicated to

Mildred Snider

a mom who provided the best foundation for life

of which any daughter could be blessed.

To God be the glory for my mother and all of my blessings.

ACKNOWLEDGMENTS

With pleasure, I extend a very special thanks to my major advisor, Professor Donald Marcotte for his most valuable guidance and advice. Tremendous appreciation and thanks are also extended to the other members of my committee, Professors Shlomo Sawilowsky, Joseph Dunbar, and Ronald Brown for their guidance, constructive criticism, valuable advice, time and consideration.

I am truly appreciative to all of my coworkers and friends who have in some way been supportive and contributed to my efforts. I am most grateful to my Director and friend, Julia Simmons, whose encouragement and valuable advice provided a stabilizing factor for the completion of this study. My thanks to Sandy Driscoll and members of her staff for their gracious assistance with the data collection.

A very, very special "hug" to a dear friend, Sandy Snedecor, for her very patient assistance and caring support when ever needed. True friends really do share late evenings, cold days, and sometimes incredibly morbid jokes that only they can appreciate.

My deepest gratitude and appreciation to my family. My husband, Roy, and my son Christopher, for providing love in my life. Thanks to my husband for never losing faith in me and enabling me to pick up the pieces after the lost of my Mother. His devotion helped me to persevere and believe once again in myself. Thanks also to "Uncle" Charles who in his unique manner has always "been there."

Finally, my everlasting gratitude to my late mother, Mildred Snider, who provided the best foundation for life of which any daughter could ever be blessed. To God be the glory for my Mother, my family, and all of my blessings.

Table of Contents

List of Tables

List of Figures

PREFACE

Standardized admission test scores such as the Scholastic Achievement Test (SAT) and American College Test (ACT) for college admission and the Medical College Admission Test (MCAT) for admission to most medical schools can impact this country educationally, economically, and culturally. At the secondary level, SAT and ACT results have evolved into a barometer for evaluating the overall academic quality of many school systems and teaching effectiveness within school districts. The economic influence of these test scores ranges from state funding to real estate values. In the selection process for admission to college and medical school, weight given to standardized test scores impacts the diversity of the student population, in terms of gender, race, and socioeconomic status composition of the student body.

The primary purpose of this study was to determine the effectiveness of test preparation (coaching and test-wiseness) on the MCAT. Also of interest was the combined predictive effect of test preparation and selected demographic variables on MCAT scores.

This study is organized into five chapters. Chapter I, Introduction, discusses the admission process (purpose and criteria) and the concept of test preparation techniques (coaching and test-wiseness); purpose of the study, relevant definitions, and limitations. Chapter II, Review of the Literature, provides a historical discussion of various aspects of test preparation effectiveness related to standardized admission tests. Chapter III, Methodology, presents the study population, research design, statistics used to analyze the data, assumptions of the study, and research questions. Chapter IV, Results, provides an analysis of the results per research question purposed. Finally, Chapter V, Discussion and

Conclusions, summarizes the study, discusses the findings, the implications, and provides recommendations for future related research.

The overall results of this study may not have provided a conclusive empirical paradigm supporting the effectiveness of test preparation for the MCAT; however, aspects of the findings could serve to contribute to evidence of the importance of test-taking skills, test-wiseness, and of need for constructively initiating the test preparation process at an early stage. The analyses of data and the conclusions of the study provide implications for academic advisors, counselors, teachers and parents in their capacities regarding examinees and standardized admission test preparation.

CHAPTER 1

INTRODUCTION

The proceeding section discusses the relevance of standardized aptitude test scores in the admission process for medical and postsecondary schools. The basic goals of medical school admission committees are presented. A description of the typical medical school admission committee also is presented in context with its use of standardized admission test scores. The perceived effectiveness of test preparation (coaching and test-wiseness) programs for admission tests by medical school candidates and college applicants are examined relative to the admission process.

Medical School Admission: Purpose and Criteria

The admission process to medical school is designed to assess the ability of the applicants to successfully complete a medical education curriculum, retain a critical knowledge base, and subsequently practice medicine. The ultimate goal and major mission of medical schools are to provide competent clinicians who are well trained scientifically, as well as in the humanists aspects of health care (AAMC Admission Requirements, 1997; Barzansky, 1993; McGaghie, 1990; Spooner, 1990). McGaghie (1990) defined medical competency as the integration of the technical knowledge of bioscience and clinical medicine with values of human qualities such as character and integrity.

Medical school admission committees vary in the criteria used for the selection of prospective students. Typically, the selection process aspires to exemplify the ethos of the mission of the institution. Each committee is unique per the mission and objectives of their respective schools in seeking evidence of personal qualities and attributes desirable

in physicians, as well as potential to exemplify goals of the school (AAMC, 1997; Edwards et al., 1990; McGaghie, 1990; Mitchell, 1990). Most committees use traditional predictors of college grades (GPA) and scores from the Medical College Admission Test (MCAT) to assess a candidate's ability to achieve academic success (Jones, 1986; Mitchell, 1990). While these predictors may lose their relationship during the latter phases of the clinical training, "many admission committee members place an inordinately high value on academic and test credentials" (Spooner, 1990 p.184).

A majority of the 125 accredited medical schools in the U.S. and Canada require the MCAT for admission consideration (AAMC, 1997; Mitchell, 1990). This test is administered twice yearly by the MCAT Program Office for the American Association of Medical Colleges. As medical schools admit relatively few students from the vast numbers of yearly applicants, performing well on the MCAT is critical and becomes a "high stakes" competitive process for the applicant. Of the 43,020 medical school applicants in 1997, 16,165 were accepted. In the 1998-99 17,379 students were accepted in the entering class from a field of 41,004 applicants (FACTS-AAMC,1997; Medical School Admission Requirements United States and Canada 2000 -2001). The serious student prepares well in advance to test for the MCAT. The test often is taken more than once and well organized test preparation programs are used to enhance a successful outcome (Zebala & Jones, 1989; Koenig & Leger, 1997).

The MCAT in the Admission Process

The MCAT is a standardized aptitude test used by admission committees to predict an applicant's potential for academic success in medical school. Anderson (1990) noted the selection process for medical school applicants has become increasingly reliant

on standardized tests and at the almost exclusive usage of empirical evidence as selection criteria. A sequential test trilogy consisting of the Scholastic Aptitude Test (SAT), the MCAT, and the National Board of Medical Examiner (NBME) is dominating premedical and medical education to the extent of driving the curriculum (Anderson, 1990; Linn, 1990).

Well organized test preparation programs, referred to as "coaching programs," have been developed specifically to assist examinees in increasing their test scores. Some colleges are allowing credit for MCAT preparation courses, while test taking techniques and coaching specifically for the SAT are now considered essential course offerings at some high schools (Anderson, 1990; Moss, 1995). As gaining entrance into medical school is heavily reliant upon high MCAT scores, a majority of medical school applicants participate in coaching programs designed to improve MCAT test performance.

In a national survey, admission officers were asked to rank their respective preadmission variables as high, medium, and low in the order of importance in their selection criteria. Mitchell (1990) reported the MCAT ranked in the high category along with college GPA, the interview, and the selectivity level of the undergraduate college. Most of these listed variables are quantifiable. An index of selectivity of a college often is the average SAT and/or ACT test scores of the recent freshmen classes (Cass & Birnbaum, 1983). Although medical school literature describes a laudable focus of seeking evidence of moral virtues, the selection procedures and highest ranked criteria fail to support that claim. The range of acceptable MCAT scores varies from school to school, as does the method of factoring the MCAT score into the selection process for predicting an applicant's potential for progress in the school's curriculum (AAMC, 1997;

Mitchell, 1990).

Postsecondary Admission

As with the reliance on standardized tests for admission to medical school, college entrance exams also are weighed heavily in the decision-making process for admitting students at the undergraduate level (Linn, 1990). The criteria most commonly used in the admission process at the college level are: GPA, SAT-Verbal (SAT-V), SAT-Math (SAT-M), high school curriculum, recommendations, essay, extra curricula activities, and intangibles (Sturgeon, 1994).

The Scholastic Aptitude Test (SAT) and American College Test (ACT) scores frequently determine whether a student is accepted to the school of his/her choice. "A student's future may very well depend upon a single standardized test score" (Seaton, 1990 p. 1). Student performances on college entrance exams almost have evolved into a criterion by which the perceived quality of a class, school, or district may be measured at the community level (Smyth, 1989; Becker, 1990; Cole, 1994). Sturgeon (1994) noted that the SAT has had a tendency to take on a life of its own as it has become "the common currency" in representing academic quality (p. 7).

Test Preparation for Standardized Admission Tests

The coaching concept

To increase performance on standardized tests, many students participate in special preparation programs to help improve their test scores. These programs are referred to as "coaching' programs. In general, such programs are designed to familiarize students with the test format; improve test-taking skills and general test-wiseness; provide instruction in the content domain, and provide drill/practice with mock tests (Anderson,

1990; Bell, 1994; Dyer, 1987; Jones, 1986; Messick, 1982). In discussing test preparation techniques and strategies, the terms of coaching and test-wiseness are used interchangeably in the literature (Anastasi, 1981; Berlinger & Cassanova, 1996; Millman, 1965; Samson, 1985; Sarnacki, 1979).

The use of coaching programs has been debated over the past 30 years (Dyer, 1987). These debates include discussions on the effectiveness of coaching in increasing student test performance, ethics and accuracy of increasing scores due to test-specific preparation, and equity of access to programs by all students (Anderson,1984; Hymel, 1991; Johnson & Wallace, 1989; Seaton,1992; Zebala & Jones, 1989). Although some school districts have test-taking skills as part of the curriculum, many do not (Carries, 1995). Recognition of the effects of practice and coaching on standardized tests (i.e., the Stanford-Binet, the Moray House Test Preparation Examination, and the SAT) suggest that test-wiseness, in part, also may be responsible for score gains (Millman et al., 1965). Increasing evidence exists that coaching for the SAT has a positive effect on test score performance, indicating that the "SAT measures test-taking ability rather than aptitude" (Teague, 1992, p. 81). Kaplan (1992), of Kaplan Testing Center, found coaching was effective for the ACT, an achievement test, as well as for the SAT. The implicit benefits of participating in coaching programs to improve test scores and enhance chances of admission to a school of choice, suggest a trend that is expected to continue in the education system (Becker, 1990; Seaton, 1992; Teague, 1992).

Coaching and test-wiseness

The concepts, practices and components of test-wiseness overlap with practices and components of coaching. The terms are often used synonymously (Messick,1982;

Milliman et al. , 1965).

Since Thorndike's introduction of the test-wiseness concept as a source of variance on educational tests, psychometricians and educators have voiced mixed concerns regarding the educational merit of these tests. Although most agree that variance due to test-wiseness attenuates test validity, not all view the concept of test-wiseness as undesirable. Research has shown that a lack of test-wiseness serves to penalize the examinee, particularly with standardized tests. Therefore, test-wiseness training is advocated as the viable solution (Anastasi, 1981; Milliman, et al, 1965; Sarnacki, 1979).

Messick (1982) postulated the plausible ongoing effects of coaching involving high quality instruction in the development of knowledge and skill and/or emphasis of test-wiseness. Score gains attributed to enhanced reasoning abilities should be transferable for the aptitude construct and appropriate criteria measures. Test-wiseness that enhances score gains also can serve to attenuate construct-irrelevant test difficulty and is applicability to other testing situations. The focus of this study is to identify how students use test preparation methods and components of coaching for ultimate effectiveness to improve their performance on standardized tests.

Limitations

Limitations of this study involved some assumptions and uncontrollable circumstances. The participants were not randomly selected nor randomly assigned to groups; therefore, the study design is not a true experimental design. Other uncontrolled extraneous variables that could differentiate between the coached and uncoached groups, and within the coached group, are the use of different coaching techniques that could be related to score gains.

Information on examinees' coaching preparation experience was based on their self-report on a questionnaire completed during the registration period. Several indicators within the survey suggested whether the students filled out the questionnaire with a conscientious effort, allowing the integrity of the data to be judged as sound. The reported percentage of each coaching component (lecture, study guide, practice exam) completed in preparation for the MCAT was considered as an indicator of examinees' motivations to perform well and were not interpreted as factors due to the course itself. All participants are from the same school; therefore, the generalizability of the study to other populations require careful scrutiny.

All participants were volunteers, most of whom self-selected some method of test preparation. Self-selection bias, however, was not a concern as the issue of interest is the effectiveness of the test-preparation methods used by medical students, not the self-selection of medical students into effective and ineffective test preparation groups.

Purpose of the Study

The major purpose of this study was to identify effective coaching (test preparation) methods and components used by medical students that enhanced their test score performance on the MCAT. A second level of interest was to identify effective test-preparation components used by different sub-groups of medical students based on science GPA, gender, and ethnicity.

Survey information from incoming first year medical students regarding the type of test-preparation program(s) in which they have participated for their college entrance exams(s) (e.g., SAT &/or ACT) and for the Medical College Admission Test (MCAT) were used. Data from this survey were used to identify effective test preparation

components used by these students.

The specific purposes of the study were:

1. To determine if there is a correlation between the MCAT scores of students who participated in coaching programs for college admission (SAT and/or ACT) and also medical school admission tests and students who only participated in coaching for the medical school admission test (MCAT).
2. To determine if there is a correlation between the type of coaching method used (Self-Directed or a Formal Coarse) to prepare for the MCAT and student performance on the MCAT.
3. To determine if there is a correlation between the type of coaching component (lecture, study guide, practice exams) used to prepare for the MCAT and student performance on the MCAT.
4. To determine if there is a correlation between the amount of participation of the coaching component used (lecture, study guide, practice exams) and student performance on the MCAT.

Definition of Terms

Achievement Test: An objective examination designed to measure relevant knowledge of course specific subjects.

Coaching: Any preparatory technique or intervention procedure specifically undertaken to improve test scores, whether by improving the skills measured by the test or by improving the skills for taking the test or both. For purposes of this study, the following

are defined as:

1. Coaching method-type: a) Self-Directed: the examinee uses coaching study guides and practice exams in preparation for a standardized test in the absence of a lecturer or facilitator. b) Formal Course: the coaching program provides lecturers (in person and taped), study guides, and practice exams with evaluations and feed-back.
2. Coaching component-types: Coaching components are the a) study guide, b) lecture and c) practice exams.

Incoming first year medical students: Students who are registering for the first year class curriculum in a Midwest school of medicine.

High-Stakes Testing: Testing where the results have very important consequences for the students and/or parents, teachers, academic institutions, and academic administrators.

Standardized Aptitude Test: A test designed to predict a student's potential for success in accomplishing academic work. Aptitude tests assess from a broad domain, are not content specific, and measure the examinee's problem-solving and critical thinking skills. For purposes of

this study, the MCAT and SAT are aptitude tests.

Standardized Test: An examination designed to be administered to a large group of examinees under similar conditions. The assumption is that with all things being equal, (i.e. the same conditions, same grade level, about the same time of year) test results can be compared in a meaningful manner.

Test-Wiseness: A multifaceted concept involving test familiarization of item format and strategies for pacing, guessing, reducing anxiety, cultivating student motivation, and providing intensive instruction for developing knowledge and skills.

Transfer: The carryover and /or application of new knowledge and learned responses from one type of situation to another. In this study it is the transfer of test wiseness and test-taking skills, acquired through coaching, from one standardized test to another.

CHAPTER 2

REVIEW OF THE LITERATURE

This chapter presents a discussion of related literature regarding the effectiveness of coaching techniques and test-wise training for standardized aptitude tests employed most often as criteria for admission to institutions of higher education. The standardized aptitude tests on which this review is focused include the Scholastic Aptitude Test (SAT) and the American College Test (ACT) used for post secondary institution admission and the Medical College Admission Test (MCAT) used as part of the admission criteria by most medical schools.

The relationship of coaching to the function and use of standardized admission tests and admission requirements are discussed first. In this context, the historical perspective of the construct of coaching in educational research is presented. Next, a review of relevant education and medical education studies on the effectiveness of coaching for standardized admission tests is discussed. Lastly, qualitative issues concerning the social and economical impact of reliance on standardized tests and equity of coaching, and admission practices are reviewed in the literature.

Coaching is a term that encompasses test preparation methods specifically designed to improve examinee test scores. The primary purpose of this study was to identify coaching preparation components used by medical school students considered to be effective for enhancing their test score performance on the MCAT.

Historical Perspectives of Coaching

Relationship of coaching to standardized admission tests

A synthesis of the literature found that coaching was a controversial issue among

psychometricians, measurement specialists, teachers, and other cohorts of professional educators. Messick and Jungeblunt (1981) postulated the controversy was precipitated by a lack of understanding standardized tests (i.e., the nature of achievement verse aptitude tests) and a lack of a common understanding or consensus of the nature of coaching. These two primary issues of concern: 1) a lack of understanding of standardized tests; and 2) a lack of clarity regarding coaching, have been further confounded by the publication of methodologically flawed studies of coaching effectiveness (Becker, 1990; Messick, 1982; Powers, 1986; Seaton, 1992).

The SAT, initially implemented by the College Board in 1926, is the pioneer of college admission tests. During that era, the concept of scholastic aptitude was considered an innate, inherited, or fixed ability. Such thinking then implied that tests designed to measure aptitude were immune to effects of coaching (Dyer 1987). Erdmann (1984) noted the controversy of coaching for the MCAT to be "driven by the misconception that the MCAT is an aptitude test that measures competencies inherent in the individual and that really do not change" (p. 388).

The term "aptitude" has historic connotations of innateness (Messick,1982) which may still linger to some degree in present day society. Aptitude, in general psychological terms, is the ability to learn (Anastasi,1981; Messick,1982; Jones,1986). In an attempt to avoid the misguided perpetuation of aptitude as "fixed ability," test developers are describing and promoting aptitude test scores as indicators of "developed ability" in their publications (Johnson, 1990).

Aptitude tests were designed to measure academic ability that an individual developed over a continuum of time and experience through formal academic curricula

and in general life experiences outside of school (Anastasi,1981; Messick & Jungeblunt,1981). Aptitude tests typically are used to predict a student's academic potential for success in college (Hymes et al.,1991; Surgeon, 1994).

Achievement tests are designed to assess specific content and typically are used to make decisions regarding employment, certification, and licensure. Under certain circumstances, these tests can be cross functional as an achievement exam at one level of content may be used to predict one's performance at the next level (Anastasi,1991).

The SAT, used nationally as a college admission test for predicting a student's potential, is employed to evaluate a student's ability to apply concepts learned (Seaton, 1992). The Medical College Admission Test (MCAT) is a standardized aptitude test designed to predict an applicant's scholastic potential for achievement in medical school. This test is used by a vast majority of the 125 accredited medical schools in the U.S. and Canada (AAMC, 1998).

Given that an aptitude test measures developed ability of general intellect acquired gradually over years, it may be perceived as resistant to effects of test score enhancement in courses with limited duration (Jones,1986; Johnson,1989). This perception mediated the unfounded assumption that aptitude tests are uncoachable. Due to the nature of achievement tests, coaching effectiveness for short-term duration was readily accepted as scores on these types of tests can be enhanced through good quality instruction (Jones, 1986). Kaplan 1992 cited in Seaton (1992) found that both aptitude and achievement standardized tests could be coached effectively. Moss (1995) reported finding three published research documents studying the effectiveness of coaching for the ACT.

The distinction in design purposes does not mean nor imply that aptitude tests measure fixed intellect or ability and achievement only measures acquired cognitive knowledge. This type of misunderstanding of standardized tests contributes to, and promotes, a major part of the coaching controversy. Anastasi (1981) explained that the two types of tests may be considered as positioned along a continuum with degrees of overlap.

Jones (1986) stated that the SAT and other aptitude tests, such as the Medical College Admission (MCAT) and the Law School Admission Test (LSAT), are instruments that measure basic analytic skills that are developed from both academic and nonacademic settings over extended periods of time.

Definition of coaching

The term, "coaching," has been defined in multiple ways. Published test preparation methods have ranged from simple drill and practice exercises with sample tests or alternate forms to promote test-wiseness. Test-wiseness is multifaceted within itself involving test familiarization of item format; learning strategies for pacing, guessing, reducing anxiety, cultivating motivation; and using intensive instruction to develop knowledge and skills (Anastasi, 1981; Backer,1990; Dyer, 1987; Johnson & Wallace, 1989; Messick ,1982). "While there is no universally accepted definition for 'coaching,' the popular use of the term means training children to answer specific types of questions and providing the information required by a specific test" (Prell & Prell, 1986, p.2.)

Messick (1982) offered an all encompassing definition of coaching as "any intervention procedure specifically undertaken to improve test scores, whether by improving the skills measured by the test or by improving the skills for taking the test or

both" (p.70). This definition will be used in the present study.

Drowns, Kulik, and Kulik (1983) distinguished coaching from simple drill and practice or tutoring. Coaching strategies include test-taking skills; drill and practice using sample tests; and direct instruction of both subject content and broad cognitive skills. Drill and practice lack academic instruction, while tutoring programs are broadly focused on improving overall cognitive ability, especially in reading comprehension and math computation. Sarnacki (1976) noted test-wiseness skills are considered to be nonspecific to the test-taker's knowledge of the subject matter being evaluated and are applicable to testing constructs in general. McCormick (cited in Moss, 1995) also described the categories of test-wiseness, test familiarization, drill and practice, and content focused instruction as four components that comprised coaching.

The Medical College Admission Test (MCAT)

The Medical College Admission Test (MCAT) is a standardized aptitude test designed to predict an examinee's potential for academic success in a medical school curriculum. The first implementation of the MCAT was in 1930. The current and latest edition was administered first in April, 1991. The test consists of three multiple-choice sections and a writing sample. Two sections, Physical Sciences (physics and general chemistry) and Biological Sciences (biology and organic chemistry), are identical in format. Both test the examinee's knowledge of basic concepts and comprehension of the information presented per science domain. The Verbal Reasoning section (reading comprehension) includes questions covering a broad domain from humanities, social sciences, and natural sciences. As these questions are not content specific and are not designed to assess specific content learned in college, they evaluate critical thinking skills

(AAMC Student and Applicant Information, 1997). Overall, the MCAT is designed to elicit "higher order" cognitive skills, such as: applying concepts, critical reasoning, and problem solving.

Numerous workshops, manuals, and commercial courses are available for MCAT preparation. Zebala and Jones (1989) cautioned medical school applicants not to "treat lightly something that may mean as much as - if not more than - three years of grades combined" (p.31). While they further acknowledged the criticism of standardized tests as "culturally biased and dehumanizing;" the MCAT scores prevail as a major factor in the admission process for medical school.

The Association of American Medical Colleges (AAMC) is the sponsoring organization for the Medical College Admission Test (MCAT) and the American Medical College Application Service (AMCAS). This organization publishes three sets of student preparation materials:

Set I The MCAT Student Manual that provides test information to include format familiarization, scoring sample questions and a full-length MCAT practice test I. Strategies for studying are also provided.

Set II MCAT practice items similar to the actual MCAT test covering all four sections of the MCAT.

Set III A MCAT practice test in the form of the most recently administered test. It also provides a set of tables whereby an examinee may conduct an in-depth analysis of his/her performance identifying strengths and weaknesses.

The AAMC has conducted analyses comparing scores of MCAT examinees who prepared through commercial coaching programs, with scores of those who did not use commercial

coaching as a means of preparation. The findings indicated effective, but small gains of about one half of one scaled score point for examiness who were coached with commercial coaching programs (Jones, 1986; Zebala & Jones, 1989).

Zebala and Jones (1989) reminded examinees that the MCAT tests very basic concepts and suggested the best review is to focus on relevant topics. The authors highly recommended the use of a self-study MCAT preparation manual, such as a Complete Preparation for the MCAT by Flowers (now known as the Betz Guide). These preparation manuals are well organized and offer relevant succinct, subject reviews that correspond to the review outline in the MCAT Student Manual published by the AAMC.

Zebala and Jones (1989) recommended the Stanley Kaplan course for the availability of the hundreds of simulated MCAT exams. They explained that an examinee who scores at about 75% or 80% on the Kaplan practice exams would be expected to score in the range of 11 or 12 on the actual MCAT. Familiarity with similar questions and problems that result in reducing anxiety is considered to be the most valuable component of the Kaplan Program. For a MCAT coaching program to be effective, examinees should complete all of the practice exams, review the preparatory booklet on a regular basis, and study approximately 20 hours per week for six to nine months prior to the exam (Zebala & Jones, 1989).

Erdmann (1984) offered the premise that the coaching controversy regarding the MCAT is based on two unfounded assumptions:

1. Test score gains associated with coaching courses, due to acquired test-taking skills, are sources of variance that compromise test validity and;
2. Coaching courses provide questions from past MCATs which account for

score gains by examinee participants.

The author further stated that MCAT science subtests test achievement not aptitude; therefore the validity would be highly questionable if it were not sensitive to content focused coaching.

Shen et al. (1997) conducted a study predicting academic performances of medical students using cognitive ability and personality characteristics as criteria. Findings indicated MCAT scores were strong predictors of academic performance in medical school; however, "the predictive power dropped sharply when clinical performance and personal suitability were part of the performance evaluation" (p. 781).

Kaplan, the oldest and largest, commercial coaching program, reports preparing students for the MCAT program for more than 35 years throughout North America. "In the past fifteen years alone we've helped over 250,000 students" (Kaplan Manual, 1997-98, p. vii). Kaplan (1997) described the MCAT as assessing the examinees' "thought process", not "thought content" (p. 9).

Over 50,000 examinees complete the MCAT annually, of which one-third are repeat test takers (Koenig & Leger, 1997). Jones (cited in Koenig & Leger, 1997) reported data from the former MCAT edition (post 1991) indicating that repeaters who prepared using a commercial coaching course had higher score gains than examinees who did not use a commercial coaching course. They scored about 0.5 higher on the science, 0.25 higher on the skills analysis (SA): quantitative, and about .05 on SA: reading.

The findings of Koenig and Leger (1997) supported Jones' results. Their data indicated score gains of 0.27 associated with commercial coaching preparation for repeat MCAT examinees. Koenig and Leger stated that "Test preparation styles also deserve

more attention, certain combination of studying techniques may be associated with higher gains and may help explain the differences in retest performance" (p.S102).

Introduction of Standardized Admission Test into the American Educational System

The following synopsis of the history of the SAT and the introduction of standardized admission tests is taken from "The Big Test: The Secret of The American Meritocracy by Lemann (1999). Lemann discussed the original purpose and intent of developers of intelligence tests and academic proponents who facilitated the evolution of such tests into the American educational system as academic admission tests.

The history of the SAT and standardized testing is rooted in the development of IQ tests. The first IQ test was developed in 1905 by a French psychologist, Alfred Binet. This test was used to identify students who needed academic assistance. The Binet test was revised by Lewis Terman at Stanford University. A later adaptation of the test was used during World War I by the army to test recruits in a group format. Prior to that administration, IQ tests were administered one on one, examiner to examinee.

Carl Brigham, a Princeton psychologist, further adapted the Army Alpha IQ test for use in college admissions, thus the birth of the SAT. Brigham's adaptation of the test was named the Scholastic Aptitude Test, and initially was marketed to military academies as well as a number of Ivy League schools. Carl Brigham, the original author and developer of the SAT was a self-acclaimed "reformed" eugenicist. During the 1920s, the eugenicists' movement was concerned that the state of immigration laws would allow a dilution of the American society they perceived as racially superior. People of color were not an issue in that the acknowledgment of their humanity was none existent and therefore was of no concern. Eugenicists recognized the White race only and categorized it into

Nordics, Alpines and Mediterraneans. Brigham's analysis of the test results were published in "A Study of American Intelligence" (1923). He described Nordics as inherently intellectually superior to lower classes of Whites – Jews and Mediterraneans. Brigham later recanted his interpretation of his analysis and renounced his position on eugenics in his publication of "A Study of Error" (1932).

The official adaptation of the use of the SAT as a Harvard scholarship test by assistant deans, Henry Chauncy and Bill Bender in the 1930s served to proliferate acceptance of the test at all Ivy League schools. In 1944, the SAT was again further revised and used as an Army-Navy College Qualification Test. This test was administered to approximately 300,000 examinees, which was a substantive increase from five to ten thousand that previously had been administered on an annual basis.

Conant, president of Harvard during the introduction of the SAT, envisioned his adaptation of this test as a means of altering the power base in the country. He perceived American society to be evolving into a class-bound, aristocratic society with privileges to education based on wealth and heritage rather that ability. Those holding the power were all male, Caucasian, and protestant. Conant's use of the SAT was a mechanism to establish a natural aristocracy among Americans, a society that promoted people on the bases of intelligence and talent. The specific focus of the test was on aptitude rather than achievement or test of mastery of an educational curriculum. This focus could provide a bias to economically disadvantaged students who were not privy to high quality preparatory schools. The test was expected to help establish an elite society based on "innate" ability and talent rather than wealth and birth background. The new elite could have attended public school, come from modest backgrounds, and develop into liberal

ideologists who would want to "serve" for the good of their country.

The SAT evolved into establishing an elite group of high scoring students who could afford test preparatory courses that range from two to five years. "Every conceivable meritocracy [ruled by the best], degrades over time into an aristocracy" (LeMann, 1999). Test preparation institutions and organizations have developed into a mechanism by which the ideology of meritocracy has evolved into an aristocracy consisting of well educated, highly paid, power brokers from Ivy League and Berkley-like institutions.

The Scholastic Assessment Test (SAT)

Coaching effectiveness for standardized tests prevails as a controversial issue in part due to a lack of agreement by the researchers, educators, and measurement specialists regarding techniques and activities that comprise coaching (Becker, 1990). Scholarly investigations of the effectiveness of coaching for standardized tests, particularly the SAT, began over four decades ago. A majority of research on coaching effectiveness focused on the standardized aptitude test now referred to as the Scholastic Assessment Test (SAT), formally the Scholastic Aptitude Test (Becker, 1990; Jones, 1986; Powers, 1987). Moss (1995) offered two basic reasons for targeting the SAT: 1) it is accepted by the most competitive colleges and universities; and 2) it has evolved into a gauge for evaluating the academic quality of high schools (p. 3).

In a 1953 Dyer study, (cited in Dyer, 1987), research was conducted to determine if coaching helped increase score gains on the SAT. He concluded that perhaps it was minimally successful in increasing scores. Coaching programs initially gained national attention in the early 1950s when the SAT became a proxy for admission to colleges

(Messick, 1982; Messick & Jungebult, 1981).

The content of the SAT is designed to draw from cognitive skills that are developed over a continuum of time and experience, in both academic and nonacademic settings. Therefore, it is not limited to subject matter or curriculum (Messick & Jungblut, 1981). The College Board administered the first SAT in 1926, which has since become ingrained into American society as an academic icon (Beaver, 1996). The new SAT is designed to acknowledge the diversity of the examinee population, as well as emphasize problem-solving and critical thinking skills. The new edition consists of two major components:

SAT I: Verbal and Mathematical Reasoning Tests is an enhancement of the former SAT-M and SAT-V sections.

SAT II: Writing and Subject Tests is a new test that builds on the former Achievement Test and is designed to assess proficiency in the English language.

College Board officials reported that almost 2.8 million students at various grade levels registered for the SAT in 1997-98. This number represented about a 54% increase from the 1996-97 school year. About 90% of the four year colleges and universities require the SAT as a part of their admission process (College Board On Line, 1998). Graff (1993) described the annual administration of the SAT as "a rite of passage" (p.7).

The coachability of the SAT has been acknowledged by the College Board and the Education Testing Service (ETS) after many years of denial (Beaver, 1996; Becker, 1990; Teague, 1992). ETS claimed that 40 hours of coaching is required to effect score gains from 30 to 40 points. In contrast, Kaplan and other commercial programs claimed average score gains, due to coaching, in excess of 100 points (Beaver, 1996). The new SAT is

purported to be less susceptible to the effects of coaching in that it is measuring problem solving and critical thinking skills that are not directly from course work content.

Commercial coaching courses can cost about $600 or more per examinee. Millions of dollars are spent annually in buying test preparation material (Bracey, 1990; Teague, 1992). The administration of the SAT is reported to generate annual revenue of over $120 million (Beaver, 1996).

Studies of Coaching Effectiveness and Standardized Admission Tests

Coaching components

Numerous traditional reviews, meta-analyses, and primary research studies have been published on coaching effectiveness. Most of these studies have examined various aspects of coaching relative to program duration (total hours versus number of days) and type of school (private versus public). Most studies exhibit methodological and/or design flaws (Becker, 1990; Messick & Jungblut, 1981).

Pike (cited in Becker, 1990) developed a conceptual framework for explaining coaching from the premise of the components of an examinee's test score. He asserted that a test score consisted of three components:

1. A true score which represents developed knowledge, relevant analytical skills, and an extent of overlearning.
2. A primary and a secondary score. The primary component integrated general test-wiseness with test content and examinee ability;
3. the secondary score involved confidence and efficiency.

Coaching techniques focus primarily on the test-specific component. Pike's literature review focused on short-term and intermediate instruction for the SAT. He found

meaningful coaching effects for the SAT-Math (SAT-M).

Bond (cited in Becker, 1990) modified Pike's conceptualization and labeled the true score as alpha (α) and the test-specific score as beta (β). Both components were targeted by coaching. Test-oriented instruction is most effective on the α component; however, standardized tests are developed to minimize the effects of coaching on this component. The degree to which coaching may effect the α component depends on the amount of content relevant instruction provided.

Studies of the aspects of coaching duration

The Federal Trade Commission (FTC, 1979) conducted the first study on the effectiveness of commercial coaching programs. The programs studied included students preparing to take either the SAT or the Law School Admission Test (LSAT). While the findings of this study indicated relatively strong score gains for coached students, it was plagued with design flaws and a reanalysis of the data was later completed. (Messick, 1981).

Messick and Jungeblut (1981) reviewed studies covering a 30-year period from 1953 to 1980 to investigate coaching effects on SAT results. The focus of this review was to examine the relationship between test score gains and the length of time students committed to each type of coaching program. Study results were compared with growth from a secondary curriculum exposure that could occur in the absence of coaching. Based on the results of this review, the authors concluded that a nonlinear relationship existed with students who devoted more time to coaching thereby experiencing diminishing returns for their efforts. They concluded that a "threshold" of three hours of coaching minimally is required to produce any SAT score gains above that expected with retesting.

To achieve average score increases greater than 20 to 30 points, especially for the SAT-Verbal (SAT-V), a student would have to commit almost as much time to coaching as she/he does to full time formal schoolwork. This study was re-examined by Becker (1990) who described it as "confounded with design features which increased the apparent effect in their analysis" (p. 405).

Kulik et al. (cited in Becker, 1990) conducted a meta-analyses of three types of SAT coaching programs: short-term programs of test-taking strategies and practice; long-term programs of drill and practice on test items; and programs of instruction in broad cognitive skills. Program instruction was further examined for features of test-wiseness, anxiety reduction activities, specific domain content instruction, and practice on specific item formats. The variables were not significantly related to coaching effects for the SAT, but test-wiseness was found to be more effective than the other components.

Becker (1990) conducted a meta-analysis assessing the effects of practice on specific test items. Practice was found to be the single most important component of coaching that was associated with score gains on the SAT.

Drowns et al. (1984) reviewed 14 controlled studies and categorized then by program duration into two groups: (a) average duration (3 to 9 hours) was .8 SD units and (b) longer duration (> 9 hours) was .16 SD units. Contrary to the findings of Messick and Jungblut (1981), these results were not statistically significant and the analyses did not reveal significant differences based on duration or methodology. Care must be taken in comparing the two results as the investigation by Messick and Jungblut (1981) included several studies without controls and of longer duration.

Powers and Alderman (cited in Smyth, 1989) assessed the effectiveness of longer

term coaching and found average score gains of about 8 points with relatively test-wise students. They concluded that a threshold is reached rather quickly after which significant gains are achieved slowly in conjunction with the development of knowledge and skills. Moss (1995) explained that long-term coaching courses in content-domain instruction, targeting the α test score component, has no significant relationship to score gains on the SAT. He concluded that this coaching technique may be more appropriate for achievement tests. Scott et al. (1980) did not find a significant relationship between score gains for the National Board of Medical Examiners (NBME) and time devoted to coaching activities in a commercial coaching program. The NBME is an achievement test that medical students complete after year two of the medical curriculum.

Studies of the effects of coaching on item-types

Powers (1986) studied the differential susceptibility of item-types to coaching practice. The findings were that the combination of types and proportional mix of item-types in a test may be essentially related to the susceptibility of the test to coaching with complex formats having a higher susceptibility. Characteristics of test item-types that are related to the coachability of the test include: complexity of direction and format, unfamiliarity with computational operations, and interrelations among the answer choices.

Since the initial implementation of the SAT in 1926, over 38 distinct item-types have been used. As with the SAT, substantial item-type variations per edition exist with other standardized admission tests such as the Graduate Record Exam (GRE), the Law School Admission Test (LSAT) and the MCAT. Item-type susceptibility to coaching effect size may be significantly influenced by the combinations utilized within a test.

Quality of Studies Assessing Coaching Effectiveness

The FTC study and reanalysis of study data by Messick and Jungblut (1981) had methodology flaws, lacked control groups for some studies, and used nonequivalent control groups for others studies. All examinees who enrolled in commercial courses were volunteers that could be construed as selection bias in the study. Using nonofficial posttests in some instances presented a problem with maintaining examinee motivation. Messick & Jungblut did not disaggregate the different designs in their reanalysis of the FTC study; therefore, they could not discuss the studies separately.

Effects of coaching on validity

Validity is a concern with standardized tests if they have no related objectives of instruction (Marlow, 1994). A test is considered "valid" to the extent of the accuracy of the inference regarding the student's true comprehension. Thorndike is credited with introducing test-wiseness as a construct described as a persistent general attribute of examinees that contributes partially to test score differences among individuals (Millman et al, 1965; Samacki, 1979). Since Thorndike's thesis regarding test-wiseness as a construct and a source of variance that could negatively impact the validity of test scores, there have been arguments in support of making examinees test wise and arguments that consider the construct as unethical.

Coaching techniques that may serve to overestimate an examinee's knowledge and skills can have implications for validity attenuation. Techniques that use answer-selection tricks to increase test scores can negatively affect both the construct and predictive validity of the test (Jones, 1986; Linn, 1990; Messick, 1982).

Test preparation in test-wiseness can enhance test validity by reducing anxiety and

errors made due to nontest factors (Anastesi,1981; Millman et al, 1965; Sarnacki, 1979). Overall validity attenuation may occur to the extent of the advantage of being test-wise and coached examinees in comparison to the noncoached (Anastasi, 1981; Jones, 1986; Mehrens & Kaminski, 1989). Linn (1990) stated that "large changes due to short-term instruction in test-taking strategies that are unrelated to nontest activities or understanding would threaten the validity of an admission test regardless of its location on the aptitude-achievement continuum" (p.315).

Critics have taken issue with coaching for the SAT in terms of the construct validity of the test. Messick (1982) suggested a resolution may lie in determining the relationship between the amount of student time and effort committed to some type or types of coaching and the magnitude of associated score gains. The issue of coaching effectiveness may be further advanced when the question of "to what degree and by what means" is investigated (Messick, 1982, p. 7).

Erdmann (1984) reported results of a five-year study involving first-time Spring MCAT examinees who participated in a coaching program prior to repeating the MCAT in the Fall. Analysis indicated general score gains of approximately one half of a point in the science subtests that could be associated with coaching. The magnitude of these results did not suggest unreasonable learning from a meaningful time period (about 12 weeks) of content specific course work in the MCAT sciences. The reading and quantitative skills analysis (SA) subtests indicated increases of about .2 of a point for each. The assumptions of score gains due to highly similar or identical questions used in coaching reappearing on subsequently administered MCAT tests was rejected by data from this study. The skills-type items in general, and the reading SA specifically, should have been more affected by

test-taking techniques due to the structure (i.e., longer and more complicated items) as discussed per Powers (1986).

Jones (1986) considered content-specific instruction, a major coaching component of many programs, not to be an inhibitor of test validity; rather it serves to increase the examinee's level of knowledge and skill. Jones further noted that the MCAT science section could be expected to be affected by such coaching techniques; not to be affected would call into question the construct validity of the test. Coaching components that enhance abilities and skills through high quality instruction do not attenuate construct validity nor predictive validity of the SAT (Messick, 1982, p.81).

Mitchell (1990) reported (validity coefficients) data indicating admission variables (GPA and MCAT scores) to be much less than perfect predictors of clinical performance in medical school. Anderson (1990) cited correlations between the MCAT and clinical performance to have negative coefficients with some observations.

Powers (1985), in an analysis of coaching effects on the GRE General Test (analytical section), found that validity was negatively affected by coaching. Jones (1986) noted that

> . . . the effect of commercial coaching courses on performance on the NBME examinations or the MCAT Science Knowledge and Science Problems subtests threatens the validity of these tests only to the degree to which the magnitude of the effect exceeds reasonable expectations based on the nature and duration of the review program or to the extent to which the effect changes the predictive relationship between the test taker's performance and subsequent academic performance. (p.275)

Equity Issues of Coaching

Equity of coaching availability

Messick (1982) purported that the coaching controversy continues to involve three

primary issues: a) effectiveness; b) equity of access and; c) construct validity. He proposed that the question regarding coaching could be what type of student who would most likely benefit from coaching. Messick's findings concluded that students who were not test wise were disadvantaged in taking standardized tests. Being test wise was the coaching component found to account for the larger than average score gains. Anastasi (1981) cited studies that indicated the effectiveness of short-term test-taking strategies and practice that served to equate the level of test sophistication among educationally disadvantaged examinees. If certain test item-types are found to be more susceptible to coaching then they could be considered to be less fair to the noncoached examinee (Johnson &Wallace, 1989). In an randomized study of coaching effects involving minority and disadvantaged students for the Graduate Record Examination (GRE) aptitude test, Powers (1989) demonstrated that disparity in prior academic preparation would be reflected in between group test score differences. A focus of the study was on minority and disadvantaged students. When afforded the same quality of test preparation, the low-scoring examinees could achieve proportional score gains. Raising test scores by special preparation (coaching) techniques could be necessary due to the unequal quality of institutions, which is a serious consideration when interpreting test scores (Bell, 1994). Linn (1989) purported that performance differences could be a reflection of the quality or inequity of secondary education and preparation as well as cultural exposure and idiom. Messick (1982) queried the fairness of coaching if, at a minimum, not all students are exposed to, or made aware of, the significance of such special preparation.

CHAPTER 3

METHODOLOGY

This study focused on test preparation (coaching and test-wiseness) effectiveness. The literature suggested that performance on standardized admission tests can be enhanced through test preparation programs that use various methods and components. This research investigated the relationship between test preparation (methods and components) and Medical College Admission Test (MCAT) scores. The methods and components were selected on the basis of theoretical considerations and previous research findings that incorporated the concepts and techniques of coaching and test-wiseness. An investigation of the relationship of test preparation effectiveness on standardized admission test scores across selected demographic variables was undertaken as well. Also of interest was the combined predictive effect of test preparation and various demographic variables on MCAT scores.

Population of the Study

The population of this study consisted of 383 incoming first year medical students at a Midwest school of medicine. All students had completed the MCAT at least once. All had taken the SAT and/or the ACT, with the exception of 3 students. A majority (81%) had used some form of test-taking/coaching study in preparation for those standardized college admission tests.

Measurement

Results from a self-report questionnaire administered to 383 incoming first year medial students at a Midwest school of medicine were used for this study. Students were queried regarding their method(s) of test preparation for college admission and/or for the

MCAT. Participants were also requested to provide an estimate of the amount of the test preparation materials completed each time they participated in test preparation activities.

Test-preparation methods were categorized into two major types: Self-Directed study and Formal Course (commercial) work. The two types of methods were further categorized into components. The Self-Directed study consists of two components (study guide and practice tests). The Formal Course has three components (study guide, lecture, and practice tests). These components corresponded with the major organizational structure of self-study manuals (study guide and practice tests) and commercial courses (study guides, lectures, and practice tests).

The instrument design provided choices for student recall of the methods(s) of test preparation used and provided a range of values for the estimate of preparation components completed per method used. The survey was structured to collect responses from participants who had repeated the MCAT a minimum of four times.

The Self-Directed study items focused on study guides of Barrons and Betz. The Formal Courses listed Kaplan, Princeton and Excell commercial courses. An "other" category for the method was provided for participants who had used some method(s) not listed on the survey, such as college-based programs/workshops or seminars, class notes and/or other commercial programs. These responses were subsequently categorized as either Self-Directed or as a Formal Course.

Further data were collected from student files (i.e., gender, ethnicity, undergraduate school type, science GPA and MCAT scores). Demographic data from student files were collected in a confidential manner. Utilization of all data was facilitated in a manner to maintain student anonymity. Permission was obtained from the Associate

Dean of Academic and Student Programs and the University Human Investigation Committee. Letters of permission are found in Appendix A.

Nominal scale data in the study include test-preparation methods (Self-Directed or Formal), components (study guide, lecture, practice tests) and demographic data (undergraduate majors, ethnicity, and gender). Interval scale data involve student GPAs, MCAT scores, "age" at the time of testing for the MCAT, and percentage of test-preparation components completed.

Variable Description

Grade Point Averages: Over all Grade point averages (OGPA) are numerical means for all course work completed at the undergraduate level. Science grade point averages (SGPA) are numerical means for all science course work completed at the undergraduate level.

Undergraduate Majors: Majors were classified according to the discipline of the earned undergraduate degree, such as: biological science, physical science, premedical/health science, humanities, math/statistics, and other.

Ethnicity: The student self-description categories of ethnicity are: African-American, Asian, Caucasian, Hispanic and Native American.

MCAT Scores: Scores on the multiple-choice sections are reported as three scaled scores, each ranging from a minimum of 1 to a maximum of 15. Student scores used were same as reported in the official files from the American Medical College Application Service (AMCAS).

High School Test-preparation for the SAT and ACT: Choices were provided for "yes" or "no" responses to test-preparation methods to include self-directed study or

organized study groups or workshops sponsored by the high school, organizations outside of the high school or commercial courses.

Percentage of Test-preparation Component: A range of values (100%, 75%, 50%, 25% and < 25%) was provided for an estimate of the amount of test-preparation component(s) completed.

Operational Definitions

Achievement Test: An objective examination designed to measure relevant knowledge of course specific subjects.

Coaching: Any preparatory technique or intervention procedure specifically undertaken to improve test scores, whether by improving skills measured by the test or by improving skills for taking the test or both. For this study, the following are defined as: 1)Coaching method-type: a) Self-Directed: the examinee uses coaching study guides and practice exams in the absence of a lecturer or facilitator. b) Formal Course: the coaching program provides lecturers (in person and/or taped), study guides, and practice exams with evaluations and feed-back; and 2) Coaching component-types: Coaching components are the a) study guide, b) lecture and c) practice exams.

Incoming first year medical students: Students who are registering for the first year class curriculum in a Midwest school of medicine.

High-Stakes Testing:	Testing wherein results have very important consequences for the students and/or parents, teachers, academic institutions, and academic administrators.
Standardized Aptitude Test:	A test designed to predict a student's potential for success in accomplishing academic work. Aptitude tests assess from a broad domain, are not content specific, and measures examinees' problem-solving and critical thinking skills. For purposes of this study, aptitude is measured by the MCAT and SAT.
Standardized Test:	An examination designed to be administered to a large group of examinees under similar conditions. The assumption is that with all things being equal, i.e. the same conditions, same grade level, about the same time of year, test results can be compared in a meaningful manner.
Test-Wiseness:	A multifaceted concept involving test familiarization of item format and strategies for pacing, guessing, reducing anxiety, cultivating student motivation and providing intensive instruction for developing knowledge and skills.
Transfer:	The carryover and/or application of new knowledge and learned responses from one type of situation to

another. In this study it is the transfer of test-wiseness and test-taking skills, acquired through coaching, from one standardized test to another.

Data Analysis

The techniques used in the analysis of the data consisted of a) descriptive statistics such as mean, standard deviation and frequencies; b) correlation studies; c) t-test; d) analysis of variance; e) regression analysis and f) multivariate analysis.

The methods of analysis were chosen to determine the test preparation method and component combinations per method that affected MCAT test performance. Variable coding, where appropriate, was used for the analysis of nominal data. A summary of the variable characteristics and measures is provided in Table 1.

Table 1

Summary of Variable Characteristics and Measures

Variable	Scale Type	Coding
Gender	Nominal	1, 0
Age	Interval	18-40
Ethnicity	Nominal	1, 3
Science major	Nominal	1, 0
SGPA	Interval	0.00 – 4.00
OGPA	Interval	0.00 – 4.00
Test Preparation		
Method	Nominal	1- 2
Components	Nominal	1- 3
College Admission Test Preparation	Nominal	1 0
MCAT Scores		
Biological science	Ordinal	1 – 15
Physical science	Ordinal	1 – 15
Verbal reasoning	Ordinal	1– 15

Research Hypothesis

The following null hypotheses were investigated:

H_{01}: There is no significant difference between the MCAT scores of students who participated in test -preparation methods for college admission (SAT /ACT) and medical school admission (MCAT) tests and students who only participated in test-preparation methods for the medical school admission test (MCAT).

H_{02}: The type of test -preparation method (Self-directed or a formal organized commercial course) used to prepare for the standardized admission test will have no effect on student performance on the MCAT.

H_{03}: The type of test-preparation component (study guide, lecture, and practice tests) used to prepare for the standardized test will have no effect on student performance on the MCAT.

H_{04}: The amount of participation (percentage of preparation component completed) per method used to prepare for the standardized test will have no effect on student performance on the MCAT.

Assumptions Regarding Statistical Analyses

Assumptions that underlie some multivariate procedures and most statistical test employed for parametric analyses are normality, linearity, and homoscedasticity. Multivariate analysis of variance (MANOVA) is considered to be somewhat robust against departures from normality. A scatter plot revealed that data were moderately skewed in a negative direction. Employing medical school students in a study uses a

homogenous as well as self-selected population: therefore, the related data cannot be assumed to be normally distributed.

The square root reflex transformation was used to allow data to meet the assumptions of normality and linearity within acceptable ranges. A variable is reflexed by forming a constant (k), the largest score in the distribution plus 1, and subtracting each value in the distribution from the constant (k-x). When data have negative skewness, the recommended "best strategy" is to reflex it and then apply the appropriate transformation for positive skewness (Tabachnick & Fidell, 1996).

For multivariate analysis, the analog of homogeneity of variance for individual dependent variables is homogeneity of variance-covariance matrices. The rationale for this assumption is that variance-covariance matrices within each cell of the design are sampled from the same population variance-covariance matrix and therefore can reasonably be pooled to create a single estimate of error. The Box's M test was used to test for homogeneity of covariance.

Input.

Data from the survey questionnaire were sorted into a spreadsheet and subsequently downloaded into SPSS (Statistical Package for the Social Sciences) version 9. This file contained data from the survey and a dictionary of terms defining the data to the system.

Computational method

SPSS (version 9) was used for computational analysis. The results are presented in narrative form as well as in charts, tables, and graphs.

Research Design

A survey design was chosen. Survey research has a strong propensity for emphasizing representative samples and the conceptual facility for eliciting a large amount of quality information from a large population in a cost effective manner. Kerlinger (1986) stated that "despite its evident potential value in all behavioral research fields, survey research has not been used to any great extent where it would seem to have large theoretical and practical value: in education" (p. 386). Various test preparation methods and techniques and their effectiveness on standardized admission tests have been described in the literature (Haladyna et al., 1991; Jones, 1986; Kaplan, 1997; Mehrens & Kaminski, 1989; Messick.1981). The survey was constructed by the investigator based on published test preparation techniques from previous research and on test preparation methods employed by commercial coaching programs.

Instrument

The questionnaire was piloted with 25 students from a post baccalaureate (PB) program within the same Midwest school of medicine as the study population. These students had experienced the complete medical school application process including test preparation for the MCAT and completing the MCAT. This population of students also was comparable to the study population in terms of age, undergraduate major and average number of repeated MCAT testings. After completing the survey, PB students were interviewed regarding their responses. All interviews were conducted by the researcher. The purpose of the pilot was to establish clarity of content and meaningful sequencing of items prior to an official administration. The likelihood of reliability of responses to the instrument was increased through the clarity and comprehensiveness of the instructions.

Significance of the Study

Scores on standardized tests, such as the SAT and ACT for college entrance and the MCAT for admission to most medical schools, have a crucial impact on the education system in this country. The weight given to these scores in the selection process for admitting students to colleges and medical schools can therefore influence the student population in higher education regarding gender, race, and socioeconomic status (Koenig, 1997 : Self, 1990; Teague, 1992; Tekian, 1997; Wilmouth, 1991).

A majority of the research studies on the effectiveness of the test preparation programs (coaching) over the past 30 years has indicated a positive effect on test performance (Jones, 1986; Messick, 1981; Power, 1986). With the competitiveness of college requirements and stringent acceptance levels, pressure is being experienced by parents, students, high school administrators, and teachers for admittance of their students to their schools of choice and by admission committees for the high scoring students (Carris, 1995; Seaton, 1992; Teague, 1992).

The data from this study may have implications for the recruitment and retention of minority students by medical schools and in directing preparation courses for the United States Medical Licensing Examination (USMLE), a required standardized test at the completion of year two of the medical curriculum. The demand for coaching programs and test-taking materials is increasing annually. Hall and Kleine (1990) found that a majority of school systems across the country use test preparation materials. "Hundreds of millions of dollars are spent annually to purchase test preparation materials" (Seaton, 1990).

African Americans, Hispanics and Native American students comprise smaller

proportions of students in medical schools than the proportion represented by their respective groups in the general U. S. population. Studies over the years have also indicated a differential distribution of standardized test scores by race and gender. African Americans, on average, do not score as high on standardized admission tests as Caucasians and males tend to score higher than females (College Board, 1983 - 1997; Mitchell, 1990; Wilmouth, 1991). Low scores on the MCAT can be a major limiting factor in regard to being admitted into medical school (Spooner, 1990).

The results of this study could assist in developing a better knowledge base for advising premedical students and specifically accommodating individuals from diverse ethnic groups in their preparation efforts for taking the MCAT. In 1991, the American Association of Medical Colleges (AAMC) implemented "Project 3000 by 2000" wherein medical schools began to make a concerted effort to recruit, retain, and successfully graduate 3,000 minority medical students by the year 2000. The diversity of the applicant pool, with regard to race and ethnicity, is expected to experience a rapid increase that can subsequently reflect an increase in the cultural and socioeconomic diversity. The rationale for Project 3,000 by 2000 is based on the projected future increases in the minority population in the country. This increase is reflected in the public school system and subsequently in the forecasted need for physicians to accommodate population demands in general, as well as being able to relate culturally to the minority population (AAMC, 1997; Petersdorfth, 1990).

CHAPTER 4

RESULTS

This chapter presents the methods and results of the statistical analysis of this study. The major purpose of the study was to determine the effectiveness of test preparation (coaching and test-wiseness) on the Medical College Admission Test (MCAT). Also of interest was the combined predictive effect of test preparation and selected demographic variables on MCAT scores.

Research questions and findings are reported under four major areas:

1. methods of test-preparation;
2. components of the methods of test preparation;
3. estimated percentage of components completed per method used; and
4. academic levels (high school and college) of test preparation.

In answering each question, tables of frequencies and percentages are provided. When appropriate, descriptive statistics are reported. Multivariate analysis of variance (MANOVA) and regression analysis related to the research questions are also reported. A square root reflex transformation was performed to change the distribution of the MCAT scores and meet the assumptions of normality and linearity. Interpretation of the multivariate and regression analyses is based on the transformed variables. Hypotheses were tested using the alpha level of .05.

Population of the Study

The population of this study consisted of 383 incoming first year medical students at a Midwest school of medicine. The medical school and participants were selected based on accessibility. In addition, it is representative of an urban school population with

a comparably large number of ethnic minorities.

The ethnicity and gender of the students were obtained from their records. The two variables were crosstabulated to determine the distribution of ethnicity by gender. Table 2 presents the results of this analysis.

Table 2

Crosstabulation – Ethnicity by Gender

Ethnicity	Gender				Total	
	Female		Male			
	N	%	N	%	N	%
African American	28	16	16	8	44	11
Asian	21	12	47	22	68	18
Caucasian*	124	72	147	70	271	71
Total	173	100	210	100	383	100

*Hispanic and "Other" ethnic groups are included in Caucasian.

The participants were predominately male (n=210, 55%) and Caucasian (271, 71.0%). Asian students comprised the second largest ethnic group (n=68, 18.0%) and had the largest gender difference with twice as any males (n=47, 22%) as females (n=21, 12.0%). African Americans comprised the smallest group (n=44, 11%) of the respondents and included 28 (16%) females and 16 (8%) males. The ratio of females to males is the inverse of other ethnic groups represented in the study. The groups categorized as Hispanic and Other comprised the remaining 3% of the study population and were collapsed into Caucasian to form one group.

The age of students were compared by gender using descriptive statistics. The results of these analyses are presented in Table 3.

Table 3

Descriptive Statistics – Age by Gender and Race

Gender and Race					Range	
	Number	Mean	SD	Median	Minimum	Maximum
African American						
Female	28	23.86	5.74	21	19	37
Male	16	25.00	4.94	23	20	35
Asian						
Female	21	20.33	1.59	20	18	26
Male	47	20.06	.87	20	19	24
Caucasian*						
Female	124	21.94	3.42	21	18	39
Male	147	21.80	2.90	21	18	36

*Hispanic and "Other" Ethnic groups are included in Caucasian.

The female African American participants had a mean age of 23.86 (sd=5.74) years, with a median age of 21 years. The range of ages of these participants was from 19 to 37 years. The average age of the male African American participants was 25.00 (sd=4.94) years, with a median of 23 years. The ages of the African American males ranged from 20 to 35 years.

The Asian female participants had a mean age of 20.33 (sd=1.59) years, with a median age of 20 years. The Asian females ranged in age from 18 to 26 years. The average age of the male Asian participants was 20.06 (sd=.87) years, with a median age of 20 years. The range of ages for this group was from 19 to 24 years.

Caucasian females had a mean age of 21.94 (sd=3.42) years, with a median age of 21 years. The ages of the Caucasian females ranged from 18 to 39 years. The mean age of the male Caucasian participants was 21.80 (sd=2.90) years, with a median of 21 years. The range of ages for the Caucasian males was from 18 to 36 years.

The students' overall and science grade point averages were obtained from their

student records. Descriptive statistics were used to compare the students' overall grade point averages and their science grade point averages by gender and ethnicity. Table 4 presents the results of this analysis.

Table 4

Descriptive Statistics – Undergraduate and Science Grade Point Averages by Race and Gender

Race and Gender	Number	Mean	SD	Median	Range	
					Minimum	Maximum
Undergraduate Overall Grade Point Average						
African American						
Female	28	3.13	.41	3.14	2.35	3.71
Male	16	2.86	.41	2.92	2.28	3.49
Asian						
Female	21	3.60	.30	3.67	3.00	3.71
Male	47	3.62	.26	3.64	2.93	4.00
Caucasian*						
Female	124	3.58	.26	3.60	2.59	4.00
Male e	147	3.53	.34	3.60	2.34	4.00
Science Grade Point Average						
African American						
Female	28	3.00	.45	2.88	2.33	3.93
Male	16	2.63	.58	2.51	1.81	3.59
Asian						
Female	21	3.56	.35	3.62	2.73	4.00
Male	47	3.56	.32	3.62	2.80	4.00
Caucasian*						
Female	124	3.50	.29	3.51	2.52	4.00
Male	147	3.48	.40	3.57	2.20	4.00

*Hispanic and "Other" ethnic groups are included in Caucasian.

Overall Grade Point Averages. The African American females had a mean OGPA of 3.13 (sd=.41), with a median of 3.14. The range of OGPAs for the African American females was from 2.35 to 3.71. The African American males had an average OGPA of 2.86 (sd=.41), with a median of 2.92. Their OGPAs ranges were from 2.28 to 3.49.

range of OGPAs was from 3.00 to 3.71. The mean OGPA for Asian males was 3.62 (sd=.26), with a median of 3.64. The grade point averages ranged from 2.93 to 4.00 for this group.

The Caucasian females had a mean grade point average of 3.58 (sd=.26), with a median OGPA of 3.60. These females' OGPAs ranged from 2.59 to 4.00. The Caucasian males' OGPAs averaged 3.53 (sd=.34), with a median of 3.60. The range of their OGPAs was from 2.34 to 4.00.

Science Grade Point Average. The mean SGPA for African American females was 3.00 (sd=.45), with a median of 2.88. The SGPAs for these women ranged from 2.33 to 3.93. The African American males had an average SGPA of 2.63 (sd=.58), with a median of 2.51. The SGPAs ranged from 1.81 to 3.59.

The Asian females had a mean SGPA of 3.56 (sd=.35), with a median of 3.62. Their SGPAs ranged from 2.73 to 4.00. Asian males' SGPAs ranged from 2.80 to 4.00, with a median of 3.62. Their mean SGPA was 3.56 (sd=.32).

The Caucasian females had SGPAs that ranged from 2.52 to 4.00, with a median of 3.51. The mean score for this group was 3.50 (sd=.29). The average SGPA for Caucasian males was 3.48 (sd=.40), with a median of 3.57. Their SGPAs ranged from 2.20 to 4.00.

The undergraduate majors were categorized into six areas:

1. Bioscience
2. Health science
3. Humanities
4. Math/statistics
5. Physical science
6. Other

The major areas of study were crosstabulated by the participants' race and gender.

Results of this analysis are presented in Table 5.

Table 5

Crosstabulation – Major Area of Study by Race and Gender

Race and Gender	Major Area of Study											
	Biological Science		Health Science		Humanities		Math/ Statistics		Physical Science		Other	
	N	%	N	%	N	%	N	%	N	%	N	%
African American												
Female	20	71.5	3	10.7	1	3.6	0	0.0	2	7.1	2	7.1
Male	11	68.7	2	12.5	1	6.3	0	0.0	0	0.0	2	12.5
Asian												
Female	13	62.0	0	0.0	0	0.0	0	0.0	4	19.0	4	19.0
Male	36	76.5	2	4.3	0	0.0	0	0.0	7	14.9	2	4.3
Caucasian												
Female	89	71.7	10	8.1	13	10.5	1	.8	8	6.5	3	2.4
Male	88	59.9	9	6.1	10	6.8	8	5.4	26	17.7	6	4.1
Total	257	67.1	26	6.8	25	6.5	9	2.3	47	12.3	19	5.0
Female	122	70.5	13	7.5	14	8.1	1	.6	14	8.1	9	5.2
Male	135	64.3	13	6.2	11	5.2	8	3.8	33	15.7	10	4.8

Most of the participants (n=257, 67.1%) were Bioscience majors with Physical Science (n=47, 12.3%) listed as the next highest major. The remaining 79 (21%) students had undergraduate majors that included Health Science (n=26, 6.8%), Humanities (n=25, 6.5%), Math/Statistics (n=9, 2.3%) and Other (n=19, 5.0%). The same distribution trend is reflected among the ethnic groups. The undergraduate majors were also categorized into Science (Bioscience, Health Science, and Physical Science) and Nonscience (Humanities, Math/Statistics, and Other). The majority of participants (n=330, 86.2%) had majors that were classified as science with the remaining 53 (13.8%) majoring in nonscience areas.

Academic Levels (Pre/Post College and Test Preparation

Participants who had used test preparation methods at the precollege level (SAT/ACT) and at the post college level (MCAT) were identified from the data. The responses were crosstabulated by race and gender, with the results presented in Table 6.

Table 6

Crosstabulation – Received Coaching for ACT/SAT and MCAT

Race and Gender	Participated in Coaching							
	ACT/SAT				MCAT			
	Yes		No		Yes		No	
	N	%	N	%	N	%	N	%
African American								
Female	4	25.0	12	75.0	18	64.3	10	35.7
Male	12	42.9	16	57.1	12	75.0	4	25.0
Asian								
Female	13	61.9	8	38.1	13	61.9	8	38.1
Male	34	72.3	13	27.7	36	76.6	11	23.4
Caucasian								
Female	45	36.3	79	63.7	88	71.0	36	29.0
Male	55	37.4	92	62.6	110	74.8	37	25.2
Total	175	45.7	208	54.3	277	72.3	106	27.7
Female	74	42.8	99	57.2	119	68.8	54	31.2
Male	101	48.1	109	51.9	158	75.2	52	24.8

Most of the participants (n=277, 72.3%) used some type of MCAT test preparation. Fewer students (n=175, 45.7%) had participated in some type of test preparation for the high school SAT and/or ACT. The various methods included self-study books, test taking workshops, or coaching programs. As a percentage of their ethnic group, male Asians (n=34, 72.3%) were the most likely to have participated in some form of test preparation for the college admission tests (SAT/ACT), with African American females (n=4, 25.0%) least likely to have participated in test preparation. The majority of both males and females of each of the three ethnic groups had participated in test

preparation for the MCAT. Asian males (n=36, 76.6%) were the most likely to participate in test preparation, with African American females (n=18, 64.3%) the least likely to participate. Based on these findings, it appears that students were more likely to use test preparation before taking the MCAT than when taking the ACT/SAT.

Analysis of MCAT Scores and Test Preparation

The MCAT score consists of three multiple-choice sections – Biological Sciences (BS), Physical Sciences (PS), and Verbal Reasoning (VR). Each section was examined for the effects of test preparation.

Hypothesis one:

Hypothesis one stated that there was no significant difference between the MCAT scores (BS, PS, and VR) of students who participated in test preparation methods for both the college admission (SAT and/or ACT) and the professional school admission (MCAT) tests and students who had only had test preparation method for the Medical School admission test (MCAT).

Table 7 provides a description of the means for MCAT scores per type of standardized test (SAT, ACT and/or MCAT). The population of N=377 included all participants for whom MCAT scores were available. Science subtests scores were higher for examinees who had MCAT test preparation compared to those with no test preparation.

Table 7

Means and Standard Deviations for MCAT Scores per Type of Test (SAT/ACT and/or MCAT)

SAT/ACT	MCAT	N	MBS		MPS		MVR	
			M	SD	M	SD	M	SD
No	No	76	8.97	1.98	8.82	2.27	8.51	1.89
	Yes	131	9.76	1.65	9.20	1.97	9.01	1.85
	Total	207	9.47	1.82	9.06	2.09	8.83	1.88
Yes	No	28	8.79	2.10	8.46	2.27	8.79	1.97
	Yes	142	9.40	2.06	9.17	1.20	8.74	1.85
	Total	170	9.30	2.07	9.05	2.05	8.75	1.87

The scores on the MCAT for Biological sciences, Physical sciences, and Verbal Reasoning were used as the dependent variables in a 2 x 2 multivariate analysis of variance, with test participation for the ACT/SAT and MCAT used as the independent variables. The results of the MANOVA are presented in Table 8.

Table 8

Multivariate Analysis of Variance – Effects of College Admission Preparation (SAT/ACT) and MCAT Preparation on MCAT Scores for Biological Sciences (BS), Physical Sciences (PS), and Verbal Reasoning (VR)

Source of Variation	Sum of Squares	DF	Mean Square	F	P
SAT/ACT Prep					
BS	.19	1	.187	1.17	.281
PS	.14	1	.144	.62	.432
VR	.01	1	.007	.01	.942
MCAT Prep					
BS	1.34	1	1.34	8.35	.004
PS	.78	1	.78	3.38	.067
VR	.15	1	.15	1.01	.315
SAT/ACT X MCAT					
BS	.01	1	.01	.06	.810
PS	.01	1	.01	.41	.525
VR	.21	1	.21	1.45	.229
Residual					
BS	59.76	373	.16		
PS	86.57	373	.23		
VR	54.28	373	.15		
Total					
BS	61.46	376			
PS	87.36	376			
VR	54.87	376			

Note:
SAT/ACT Preparation: Roy's Largest Root = .004, p=.720, power=.140
MCAT Preparation: Roy's Largest Root = .002, p=.041, power=.670
Interaction: Roy's Largest Root = .008, p=.416, power=.260

One main effect, MCAT preparation produced a Roy's largest root of .002, which was statistically significant at an alpha level of .05. The other main effect, ACT/SAT preparation and the interaction between ACT/SAT and MCAT were not statistically significant at an alpha level of .05. An examination of the univariate F tests for MCAT preparation provided evidence of a statistically significant difference on scores for the Biological Sciences section on the MCAT. To determine how the groups were differing, the mean scores were inspected. Participants who had completed test preparation for the

MCAT (m=9.76, sd=1.65) had higher mean scores on this analysis than participants without test preparation (m=8.97, sd=1.98). The remaining tests were not statistically significant. There is mixed and partial support for this hypothesis.

Analysis of MCAT Scores and Test Preparation Method Types

Methods of test preparation for medical school were categorized into two major types, Self-directed (independent study) and Formal (commercial) Course work. For the purposes of this study, Self-Directed test preparation focused on the MCAT study guides of Barrons and Betz. The category of Formal Course focused on the commercial courses presented by Kaplan, Princeton, and Excell. Each section of the MCAT scores (BS, PS, and VR) was examined for the effects of test preparation by type. Table 9 presents the crosstabulation of the type of MCAT preparation by gender.

Table 9

Crosstabulation – Type of MCAT Preparation by Gender

Method	Gender				Total*	
	Female		Male			
	N	%	N	%	N	%
Self-directed	22	18.5	25	15.8	47	17.0
Formal Course	71	59.7	114	72.2	185	66.8
Both	26	21.8	19	12.0	45	16.2
Total	119	100.0	158	100.0	277	100.0

*Note: Students included in this subpopulation had test preparation for MCAT

The majority of the students (n=185, 66.8%) had participated in formal courses, with 47 (17.0%) reporting they had self-directed their MCAT test preparation. Forty-five (16.2%) had used both methods to prepare for the MCAT. Males (n=114, 72.2%) were more likely to participate in formal courses, as were females (n=71, 59.7%).

Hypothesis two:

Hypothesis two stated that the type of test preparation method (Self-directed or Formal Course) used to prepare for the standardized admission test would have no effect on student performance on the MCAT.

Table 10 presents the results of this analysis. The mean MCAT scores for Biological Sciences, Physical Sciences, and Verbal Reasoning by type of test preparation (Self-Directed or Formal Course) were obtained.

Table 10

Descriptive Statistics – MCAT Scores by Test Preparation Methods

Test Preparation Method	Number	MCAT Scores					
		Biological Science		Physical Science		Verbal Reasoning	
		Mean	SD	Mean	SD	Mean	SD
None	104	8.92	2.00	8.72	2.27	8.59	1.91
Self-directed	47	8.99	2.22	8.81	2.22	8.72	2.17
Formal Course	183	9.78	1.75	9.33	1.90	8.93	1.72
Both	43	9.35	1.90	8.98	2.01	8.79	2.07

The mean scores for each of the three areas tested on the MCAT were used as the dependent variables in a one-way multivariate analysis of variance. The type of test preparation method was used as the independent variable in this analysis. Table 11 presents the results of this analysis.

Table 11

Multivariate Analysis of Variance – Effect of Test Preparation Method Types on MCAT Scores

Source of Variation	Sum of Squares	DF	Mean Square	F	P
Biological Science	2.43	3	.81	5.11	.002
Residual	59.04	373	.16		
Total	61.46	376			
Tukey's a posteriori tests No test preparation < Formal Only					
Physical Science	1.26	3	.42	1.82	.143
Residual	86.10	373	.23		
Total	87.36	376			
Verbal Reasoning	.30	3	.10	.67	.569
Residual	54.57	373	.15		
Total	54.87	376			
Note: Roy's Largest Root = .042, p=.002, power=..924					

The results of the one-way MANOVA provided evidence of statistical significance for the test preparation types Roy's Largest Root =.042, p< .05. To determine which of the three sections of the MCAT were contributing to the significant finding, the univariate F tests were examined. A statistically significant difference was obtained for the Biological Sciences section of the MCAT, F (3, 373)=5.11, p=.002). The univariate F tests for the Physical Sciences and Verbal Reasoning sections of the MCAT test were not significant at an alpha level of .05. To determine which of the test preparation methods were contributing to the significant difference on Biological Sciences, all possible pairwise comparisons were made using Tukey's a posterior test procedures. Students who used the Formal Course method type of preparation (m=9.78, sd=1.75) had significantly higher mean scores than those who used no preparation (m=8.92, sd=2.00). The remaining pairwise comparisons were not statistically

significant. Hypothesis two has mixed and partial support.

Analysis of MCAT Score and Test Preparation Components

The two test preparation methods were further categorized for a more detailed description of the test preparation activity involved and to reflect the organizational structure of the self-study manuals and commercial courses. The Self Directed method has two components (study guide and practice tests). The Formal Course has three components (study guide, lecture, and practice tests).

Hypothesis three:

Hypothesis three stated that the type of test preparation component (study guide, lecture and practice tests) used to prepare for the standardized test will have no effect on student performance on the MCAT.

Many (43%) of the participants used study guides, lectures, and practice tests to some extent when preparing to take the MCAT. Table 12 provides the results of the crosstabulations of the types of test preparation and use of test preparation components that were used to summarize the findings.

Table 12

Crosstabulations – Test Preparation Component by Type of Test Preparation

Test Preparation Component	Method of Test Preparation									
	None		Self-directed		Formal Course		Both		Total	
	N	%	N	%	N	%	N	%	N	%
Study guide only	0	0.0	9	19.1	3	1.6	1	2.2	13	3.4
Lecture only	0	0.0	0	0.0	4	2.2	0	0.0	4	1.0
Practice test only	0	0.0	4	8.5	19	10.3	0	0.0	23	6.0
Study guide and lecture	0	0.0	0	0.0	5	2.7	1	2.2	6	1.6
Study guide and practice tests	0	0.0	34	72.4	19	10.3	13	28.9	66	17.2
Practice test and lecture	0	0.0	0	0.0	2	1.1	0	0.0	2	0.5
Study guide, lecture, and practice tests	0	0.0	0	0.0	133	71.8	30	66.7	163	42.6
None	106	100.0	0	0.0	0	0.0	0	0.0	106	27.7
Total	106	100.0	47	100.0	185	100.0	45	100.0	383	100.0

Students generally used a combination of test preparation components when studying for the MCAT. Use of the three components (study guide, lecture, and practice tests) together was indicated by 163 (42.6%) of the participants who took either formal courses (n=133, 71.8%) or both self-directed and formal courses (n=30, 66.7%). A combination of practice tests and study guides was indicated by 66 (17.2%) of the students. One hundred six (27.7%) of the participants did not participate in any type of test preparation and used none of the components.

Participant responses were coded for each preparation type and combination per method used (Self Directed – study guide and practice tests; Formal – study guide, lecture, and practice tests). Table 13 provides the descriptive statistics used to summarize

MCAT scores for examinees by test preparation component type and for examinees who reported no test preparation.

Table 13

Descriptive Statistics – MCAT Scores by Test Preparation Component

Test Preparation Components	Number	MCAT Scores					
		Biological Science		Physical Science		Verbal Reasoning	
		Mean	SD	Mean	SD	Mean	SD
No Preparation	107	8.92	2.00	8.72	227	8.59	1.91
All Self-directed	47	8.98	2.22	8.81	2.22	8.72	2.17
All components (self-directed and formal)	28	9.25	1.58	9.11	1.91	9.00	1.80
All study/practice (self-directed and formal)	14	9.43	2.51	8.64	2.27	8.50	2.56
All formal	139	9.78	1.79	9.37	1.89	8.95	1.73
Formal (except lecture)	40	9.90	1.57	9.37	1.92	8.97	1.63

Note: N=372, the number (268) of participants who reported specific test preparation components used.

Two hundred sixty-eight participants provided responses to the type of preparation component (s) used. An examination of the descriptive statistics showed that formal test preparation without lecture (m=9.90, sd=1.57) had a higher mean MCAT score than either all formal (m=9.78, sd=1.79) or self-directed (m=8.98, sd=2.22) for the Biological Sciences of the MCAT. When the mean scores for Physical Sciences were compared, students who had formal test preparation (m=9.37, sd=1.89) and components of formal test preparation without the lecture (m=9.37, sd=1.92) scored the same, while those who had used a combination of self-directed and formal (m=8.64, sd=2.27) had the lowest scores. Students who used all components of self-directed and formal (m=9.00 (sd=1.80) had the highest scores for the Verbal Reasoning subtest and students who used

study guides and practice (m=8.50, sd=2.56) had the lowest scores.

Scores on the three sections of the MCAT were used as the dependent variables in a one-way multivariate analysis of variance. The type of test preparation components used to prepare for the MCAT were used as the independent variables. The results of this analysis are presented in Table 14.

Table 14

Multivariate Analysis of Variance – Effect of Test Preparation Method Types on MCAT Scores

Source of Variation	Sum of Squares	DF	Mean Square	F	P
Biological Science	2.64	5	.53	3.32	.006
Residual	58.15	366	.16		
Total	60.79	371			
Tukey's a posteriori tests					
All Formal > No Preparation					
Physical Science	1.58	5	.32	1.36	.239
Residual	84.93	366	.23		
Total	86.51	371			
Verbal Reasoning	.38	5	.08	.52	.764
Residual	53.78	366	.15		
Total	54.16	371			
Note:					
Roy's Largest Root = .046, p=.005, power=..903					

Overall significance was found for test preparation components, Roy's Largest Root =.046, p=.005. When the univariate F tests were examined, statistically significant differences were found for F (5,366)=3.321, p=.006). The univariate tests for Physical Sciences or Verbal Reasoning sections of the MCAT score were not statistically significant.

To further examine the differences for the Biological Sciences significant result, Tukey's post hoc tests were used to compare all possible pairwise comparisons. The

results of this test indicated that students using the "All Formal" method of test preparation performed better than students who had no preparation. Based on the mixed findings of this analysis, hypothesis three has partial support.

Percentage of Test Materials Completed per Preparation Component

Participants were provided a range of values (<25%, 25%,50%,75%,100%) to provide an estimate of the amount of test preparation components (study guide, lecture, and practice tests) completed per test preparation method (Self Directed and/or Formal Course) they used prior to completing the MCAT.

Hypothesis four:

Hypothesis four stated that the amount (percentage) of participation per preparation component completed, per method used, to prepare for standardized tests will have no effect on student performance on the MCAT. The percentages were coded 0 thru 4 and labeled as follows: < 25%, 25%, 50%, 75%, and 100%. Table 15 presents the mean MCAT scores per the percentages of test preparation material completed by the participants.

Table 15

Descriptive Statistics – MCAT Scores by Percentage of Test Preparation Components Completed

Percentage of Test Preparation Components Completed	Number	MCAT Scores					
		Biological Science		Physical Science		Verbal Reasoning	
		Mean	SD	Mean	SD	Mean	SD
< 25%	207	9.11	2.06	8.79	2.19	8.65	1.91
25% to 49%	70	9.93	1.69	9.41	1.95	9.20	1.69
50% to 74%	47	9.87	1.31	9.47	1.61	8.94	1.83
75% to 99%	48	9.33	2.03	9.12	1.98	8.69	1.94
100%	5	9.80	1.64	10.40	1.82	8.60	2.07

The mean scores for Biological Sciences were lowest for those who completed less than 25% of the test components and highest for those who completed from 25% to 49%. The highest scores on the Physical Sciences section of the MCAT were obtained for those who completed 100% of the test preparation components. The participants who completed from 25% to 49% of the Verbal Reasoning test preparation components had achieved the highest scores on this section of the MCAT.

A MANOVA was completed using the MCAT scores for Biological Sciences, Physical Sciences, and Verbal Reasoning as the dependent variables. The independent variable was the percentage of test preparation components that were completed. The results of this analysis are presented in Table 16.

Table 16

Multivariate Analysis of Variance – Effect of Completed Test Preparation Component Percentages on MCAT Scores

Source of Variation	Sum of Squares	DF	Mean Square	F	P
Biological Science	1.84	4	.46	2.87	.023
Residual	59.62	372	.16		
Total	61.46	376			
Tukey's a post hoc tests					
25% to 49% >Less than 25%					
Physical Science	1.85	4	.46	2.01	.093
Residual	85.51	372	.23		
Total	87.36	376			
Verbal Reasoning	.74	4	.18	1.27	.283
Residual	54.13	372	.15		
Total	54.87	376			
Note:					
Roy's Largest Root = .033, p=.017, power=..806					

Overall significance was found for the test preparation component percentages, Roy's Largest Root =.033, p=.017. The univariate F tests were examined to determine which of the dependent variables was contributing to the significant results on the MANOVA. A statistically significant difference was found for Biological Sciences (F=2.87, p=.023) at the univariate level. When the univariate F test results for Physical Sciences and Verbal Reasoning were examined, no statistically significant differences were found. To further investigate the significant results for the Biological subtest, all possible pairwise comparisons were made using Tukey's post hoc tests. The results of these analyses indicated that students who completed from 25% to 49% of the preparation method chosen, performed better on the Biological Sciences section of the MCAT than those who had no preparation. Per these results, this hypothesis has mixed and partial support.

Relationship of Selected Demographic Variables to Test Preparation

Regression analysis was used to develop a model to determine which of the predictors could be used to predict MCAT test scores. Stepwise multiple linear regression analysis was used to build the model starting with all potential predictors. In a backward stepwise fashion each variable deemed not significant using the p value criteria was eliminated and a new model calculated and evaluated.

Correlation coefficients were obtained for all variables in the study. Predictor variables for the regression analysis were then selected based on the significance of the correlation and the rationale of their use. A total score for MCAT was obtained by summing the raw scores of the three MCAT subtests (Biological Sciences, Physical Sciences, and Verbal Reasoning) and calculated for use as the criterion variable in the regression model. "Age" was the participant's age at the time of testing for the MCAT. "Coached for the MCAT" was coded as a "yes" or "no" and "Test preparation component" were types of method components and component combinations that could be used in preparing for the MCAT. Table 17 presents results of the stepwise multiple linear regression analysis.

Table 17

Predictor of Total MCAT Score

Predictor Variable	β	t	Sig of t
Age	.01	2.99	.003
Coached for MCAT	-.25	-2.69	.007
Science GPA	-.58	-6.82	<.001
Test Preparation Component	-.01	-2.10	.036
Constant	4.98		
Multiple R^2	.20		
Adjusted R^2	.19		

The four-predictor variables that entered the stepwise multiple linear regression equation explained 19% of the variance in MCAT scores. A comparison of the Beta weights indicated that science GPA β=-.58, t=-6.82, p<.001 was the strongest predictor of MCAT scores, with "being coached for the MCAT" β=-.25, t=-2.69, p=.007 the next strongest predictor. Test preparation component β=-.01, t=-2.1, p=.036 was a statistically significant predictor, as was age β=.01, t=-2.99, p=.003. Using the β weights from Table 17, a regression equation using standardized scores can be formulated to generate a predicted MCAT score.

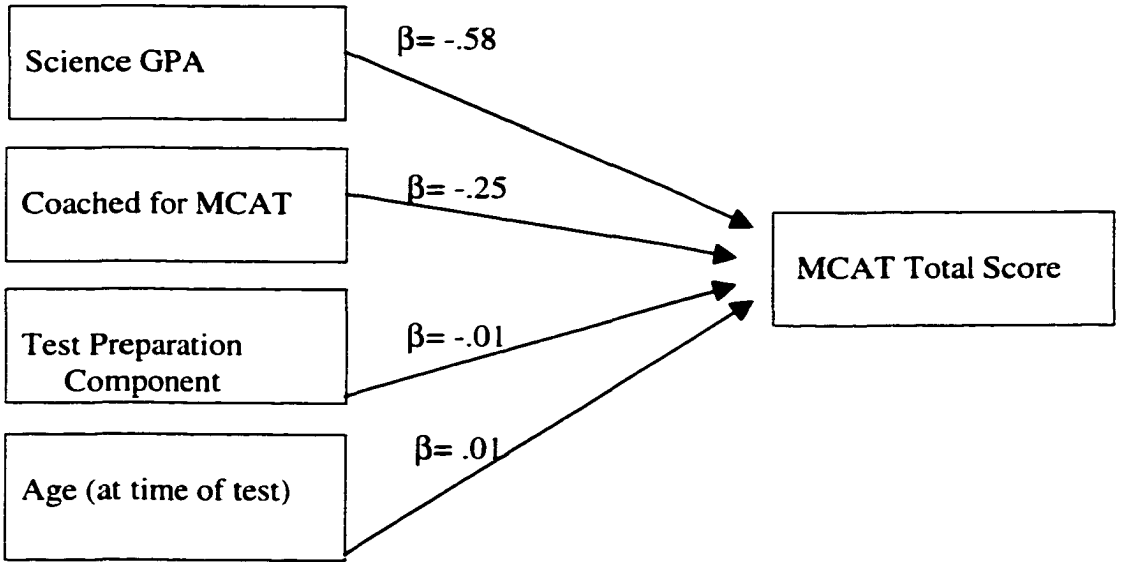
$$MCAT' = (-.58)S_Z + (-.25)C_Z + (-.01)T_Z + (.01)A_Z$$

where *S*=Science GPA; *C*=Coached for MCAT; *T*=test preparation; and *A*=Age at time of testing.

Using the Beta weights, a model was developed to demonstrate the relationship between the predictor variables and total MCAT score. Direction of the effect arrows is based on a study by Yoder (1997).

Figure 1

Model of the relationship of demographic, academic, and test preparation variables with MCAT scores



The distribution was negatively skewed to a moderate degree; therefore, a square root reflex transformation was used to "reflect" the DV variables (MCAT scores). The interpretation of reflected variables is the opposite of what it was prior to the transformation (i.e. if large and/or positive numbers meant favorable things prior to the transformation, then large and/or positive scores mean non-favorable things following the transformation). In this analysis, "age" is interpreted as negative and the three other variables are considered to be positive predictors.

CHAPTER 5

DISCUSSION AND CONCLUSION

This chapter presents a summary of the study and a discussion of the findings and conclusion. Implications based on the findings and conclusions are identified with subsequent recommendations for further research.

Summary of the Study

The primary purpose of this study was to identify effective test preparation methods and components used by medical students that enhanced their test score performance on the Medical College Admission Test (MCAT). The study was organized to examine four aspects of test preparation:

1. methods of test preparation;
2. components of the methods of test preparation;
3. estimates of the percentage of components completed per method used; and
4. academic levels (high school/ college) of test preparation.

The precollege admission tests were the SAT and ACT. The postcollege admission test was the Medical College Admission Test (MCAT). The MCAT is used by a majority of medical schools as a principal part of the admission criteria. The MCAT consists of the following four subtests: a) Biological Sciences (BS), b) Physical Sciences (PS), c) Verbal Reasoning (VR) and d) a written essay. Three subtests (BS, PS, and VR) use a multiple-choice response format. Scores for each of the multiple-choice subtests are reported on a scale ranging from 1 to 15. Test development is sponsored by the Association of American Medical Colleges (AAMC) and is administered by the American College of Testing Program. The test developer describes the MCAT as a standardized aptitude test.

However, investigators of the MCAT agree that science knowledge and problem solving subtests (BS and PS) are achievement tests (Erdman, 1984; Jones, 1986; Mitchell, 1990).

The study population included medical students who were matriculating at the same Midwest school of medicine; therefore, generalization of the study to other populations would require careful scrutiny. The medical school was chosen based on accessibility and student diversity. It is located in an urban setting with an ethnically diverse student enrollment that is representative of the population at large. This medical school also has a post baccalaureate program that is committed to enhancing student test-taking skills with a specific emphasis on standardized tests. Of the 383 first year medical students who participated in the study most were male (55%) and Caucasian (71%). Eighteen percent were Asians, mostly male and 11% were African Americans, mostly female.

Data for the study were obtained using an original survey constructed by the investigator and was based on previous research (Haladyna et al., 1991; Jones, 1986; Messick, 1981) and commercial course materials (Betz, 1998; Kaplan, 1998). The survey was a self-report instrument administered to the 383 first year medical students. Participants reported on method(s), if any, of test preparation used for the SAT/ACT and for the MCAT. They also provided an estimate of the amount of test preparation materials completed per method used. Of the study population, 72% reported having had some type of test preparation for SAT, ACT and/or the MCAT. Further data were obtained from student files with permission of the Associate Dean of Academic and Student Programs.

This study was conducted due to the lack of published data on specific test taking behavior of students preparing for standardized admission tests. The literature over the

past 15 years has noted a lack of empirical data regarding specific test preparation behavior. Messick (1982) suggested the need to determine the relationship between the amount of student time committed to some type(s) of coaching and the size of associated score gains. Jones (1986) concluded that further investigations were needed on coaching, including detailed data on the amount of time coached examinees spend on test preparation and more specifically, the nature of the coaching received. "Test preparation styles also deserve more attention, certain combinations of studying techniques may be associated with higher gains and may help explain the differences in retest performance" was an observation by Koenig and Leger (1997, p. S102).

The basic objectives of this study were to determine if medical students' test preparation behavior, academic record, and selected background characteristics contributed significantly to their test performance on the MCAT. Four hypotheses were tested in this study. Although statistically significant findings were obtained for some sections of the hypotheses, univariate and post hoc analysis provided mixed and partial support.

Multivariate analysis of variance (MANOVA) procedures were used to test the hypotheses that population means for multiple dependent variables were equivalent. MANOVA is an extension of analysis of variance (ANOVA) and is designed to test differences among groups on two or more dependent variables simultaneously. Roy's largest Root statistic was used to the test the null hypotheses. For this study, examinee types of preparation methods were labeled as Self-Directed (independent study) and Formal Course (commercial coaching). Components of test preparation methods were described as study guide, lecture, and practice tests. Analyses were conducted to examine

the effect(s) of various types and components of test preparation on MCAT scores. This study also examined the relationship of test preparation across study methods and selected demographic variables.

Discussion of the Findings

Academic levels (high school/college) and test preparation

Based on the results of the overall MANOVA, test preparation at the college level, for the MCAT, was found to be effective. In examining the univariate F tests, evidence was provided for the effectiveness of test preparation on the Biological Sciences subtest of the MCAT. This finding compared favorably with other studies. Jones (1986) also found differences in scores of examinees on MCAT science subtests. Statistical significance was not found for the effectiveness of the SAT/ACT preparation or the interaction with MCAT preparation on MCAT scores. The lack of statistical significance for either the interaction between test preparation for the SAT/ACT (pre college level) and the MCAT (post college), on MCAT scores, did not support evidence of transfer of test-wiseness and test taking skills from precollege (high school) level to college level of test preparation. Participants were asked to recall their SAT/ACT scores and few respondents could remember. Therefore, insufficient numerical data were collected to analyze this section in a manner that could have been more meaningful.

Test preparation method types

A majority of participants used the formal course (commercial) test preparation method type rather than an independent (self-directed) study. In total, 277 (72%) of the study population (N=383) had some form of test preparation.

An overall statistically significant difference was found in MCAT scores of

students who used the formal (commercial) method type compared to self-directed method type of test preparation. Tukey's post hoc analysis indicated that either form of test preparation type was more effective than not having any test preparation. This finding could possibly be due to an overlap in the test-taking strategies and content specific instruction offered by both method types, which are difficult to separate. The homogeneous population, in terms of examinee motivation to prepare for admission to medical school, could also contribute to the lack of a statistically significant difference in test performance between the two test preparation method types. Conventional wisdom, however, would suggest that students who paid for a commercial course should out-perform examinees who used a self-directed method involving less structure and less financial expenditure.

The finding, in part, does concur with research by Jones (1986) who noted the trend for coached examinees to out-perform uncoached examinees on the MCAT. Examinees who are test-wise, to the extent to which they are familiar with the test format and have experience with practice tests in the same content area, are expected to out-perform examinees who are not test-wise, (Anasita, 1981; Samson, 1985; Sarnacki, 1979).

Analysis of components of test preparation and MCAT scores

Components of test preparation methods were categorized as study guide, lecture, and practice tests. Independent study manuals included two major components (study guide and practice tests). The commercial courses have three basic components (study guide, lecture, and practice tests).

Participants were provided with a range of values from which to choose for the

estimate of test preparation components (study guide, lecture, and practice tests) completed per test preparation method (self directed and/or formal course) used. Forty-three percent of the participants used, to some extent, all three forms of the test preparation components – study guide, lecture, and practice test.

The results of the univariate F tests provided evidence of a statistically significant difference in the Biological Sciences score between the group that had no preparation and the group that had preparation. Perhaps the overlap in test-taking strategies between the components of the two methods was not amenable for the analysis.

Percentage of test components completed per preparation method

The instrument provided the following percentage choices for participants: < 25%, 25% to 40%, 50% to 74%, 75% to 99%, and 100%. Participants were asked to estimate the percentage completed of each component of the test preparation method used. A statistically significant result was found on the MANOVA, with examination of the univariate F tests indicating that MCAT results differed for the Biological Sciences subtest by the amount of completion of each component. Post hoc analysis to compare all possible pairwise comparisons employing Tukey's post hoc tests detected a significant difference for Biological Sciences between components completed in the group of 25% and the < 25% groups. This finding provides support for the effectiveness of test preparation on the science subtest of the MCAT. Researchers agree that the science knowledge and science problem solving subtests (BS and PS) are achievement tests (Erdman, 1984; Jones, 1986; Mitchell, 1990). Therefore, by the nature of achievement tests, they could be subject to the influence of specific content instruction and review. "The absence of any effect on these tests from well organized study and review over time,

whether conducted in concert with a formal course or independently, would bring into question the construct validity of these tests" (Jones, 1986, p 275). Construct validity is the extent to which a test may be said to measure a theoretical trait (i.e. science comprehension). Finding a statistically significant difference for BS and PS, but not for VR was not unexpected. The Verbal Reasoning subtest is not subject-specific and therefore not designed to access specific content learned in college/school. This subtest is much less susceptible to the influence of coaching and more closely reflects the characteristics of an aptitude test. The literature also indicated that standardized aptitude tests reflect and embrace various academic and nonacademic content that are experienced over extended time by the examinee and can greatly influence test scores. These areas cannot easily be affected by course review and specific instruction in a comparably short length of time.

Relationship of selected demographic variables to test preparation

Selected variables, relevant to the study, were used in a stepwise multiple linear regression analysis to determine if a relationship existed between total scores for the three MCAT subtests (BS, PS, and VR) and test preparation variables. The total MCAT score was the sum of the three raw scores (BS, PS, and VR) transformed. The predictor variables were 1) Science GPA (*S*); 2) Coached for MCAT (*C*); 3) Test Preparation Component (*T*); and 4) Age at time of testing (*A*). These variables accounted for approximately 20% of the variance for the three MCAT subtests. The standardized equation is expressed as follows:

$$MCAT' = (-.578)S_Z + (-.248)C_Z + (-.005)T_Z + (.003)A_Z$$

This analysis indicated that examinees who have higher science GPAs tend to have

have higher MCAT scores. The subtests of BS and PS assessed science knowledge that formed the major premedical curriculum. This finding supported evidence for the effectiveness of coaching programs that basically served as well-organized science review courses. The findings also correspond with previous research that found positive correlations between science GPAs and MCAT scores (Jones, 1983; Mitchell, 1992).

Analysis of the Age variable indicated that the younger examinees at the time of testing were more likely to have higher MCAT scores. This finding could, in part, be addressed by the premedical curriculum. Developers of the MCAT described the science assessment in biological and physical sciences as requiring knowledge at the introductory or basic course level rather than intermediate or advanced levels. Premedical examinees, who were traditional college aged and took the MCAT shortly after completing the basic level science courses while content was current, could explain why the younger age group had higher MCAT scores. Examinees who had advanced high school science courses equivalent to basic level college courses also could account for the younger age group performing well on the MCAT.

Test preparation effectiveness was supported for the achievement form of the MCAT. Examinees who have had well organized test preparation courses and/or test-wiseness training out performed examinees who lacked test taking skills and were less familiar with the test format. This finding serves to explain positive contributions of test preparation to examinee MCAT performance.

Conclusions

In summary, this study provided some confirmation of the position that test preparation, whether commercial coaching or self-directed, can be effective for enhancing

performance on the MCAT, a standardized admission test. The following conclusions were developed from the analyses of data:

1. Test preparation, whether from self-directed study or through commercial programs, has a positive affect on MCAT scores. Test preparation effectiveness was demonstrated by examining the mean MCAT scores between examinees who reported having test preparation and those who reported having no test preparation. The effectiveness of test preparation was also demonstrated using pos hoc comparisons between MCAT scores of coached and noncoached examinees.
2. Test preparation, especially commercial coaching, for the MCAT was more effective for the Biological Sciences knowledge and problem solving subtest. Coached examinees out-performed the noncoached on the science subtests of the MCAT. Students who take the MCAT soon after completion of the basic science classes tend to have higher mean MCAT scores. Age was negatively correlated with examinee MCAT scores indicating that younger examinees were more likely to attain higher MCAT scores. It was postulated that participants who were in the traditional college age group (18 –20) tested for the MCAT shortly after completing the basic science classes measured by the MCAT.

3. The Verbal Reasoning subtest of the MCAT is not as susceptible to test preparation and coaching as science subtests. Statistical significance was not found for Verbal Reasoning between the coached and noncoached examinees for any of the related hypotheses.

Implications

The analysis of data and the conclusions of the study provide implications for the students, counselors and academic advisors. Given that the coached examinees out-performed the noncoached examinees, it is strongly suggested that students seek and actively engage in test preparation strategies when preparing for the MCAT. Academic advisors and counselors should emphasize to students the need to adequately prepare for standardized admission tests in a timely manner prior to testing.

Using the science subtests as indicators of successful coaching effectiveness for the MCAT might serve to identify examinees of targeted groups who could benefit from enrichment programs. Ethnic and disadvantaged students who are a focus for recruitment, retention and matriculation in medical school, could possibly be well served by coaching programs. These groups are targeted for medical training in an effort to increase the number of physicians so as to reflect their general ethnic population in the nation.

Recommendations For Further Research

As a result of conducting this study, the following recommendations are offered:

1. Conduct additional studies to compare standardized admission test scores of students of various levels of test-wiseness, (high, medium, low).
2. Conduct additional studies to explore ways and means of emphasizing the need for a strong basic science curriculum especially for minority students at the secondary level.
3. Establish programs that will explore the examinees' level of motivation and discipline for self-directed learning and provide appropriate counseling.

4. Revise the data collection instrument using the following suggestions:
    a. Revise the instrument to collect additional data such as advanced high school courses taken and total credit hours in biological and physical sciences at the college level.
    b. Administer the instrument via the mail to allow adequate time for students to complete the survey accurately. Students would have adequate time to seek SAT/ACT scores from records and reflect on test preparation methods, strategies, and time devoted per method and strategy to test preparation.
    c. Provide categories for time devoted to test preparation that are more quantifiable, perhaps in units of time.

APPENDIX A

Permissions

WAYNE STATE UNIVERSITY
SCHOOL OF MEDICINE

Office of Academic and Student Programs
1206 Gordon H. Scott Hall of Basic Medical Sciences
540 East Canfield Avenue
Detroit, Michigan 48201
(313) 577-1450
(313) 577-1457 (Fax)

September 21, 1999

Ms. Edna Jackson-Gray, Research Assistant
Minority Recruitment
Wayne State University School of Medicine

Dear Ms. Jackson-Gray:

This letter is written to officially permit your access to medical student records at Wayne State University School of Medicine. Data is to be utilized only in the preparation of your dissertation and resulting publications. All information is to be recorded utilizing student identification numbers. As stated in Wayne State University Human Investigation Committee policies, published results based upon student data must not allow identification of an individual student. No information obtained during conduct of the study may be disclosed outside of the research. Thank you for your cooperation with these standards.

Sincerely,

Robert R. Frank, M.D.
Associate Dean for Academic and
Student Programs

RRF:ale

WAYNE STATE UNIVERSITY

HUMAN INVESTIGATION COMMITTEE
4201 St Antoine Boulevard - UHC-6G,
Detroit Michigan 48201
Phone: (313) 577-1628
FAX: (313) 993-7122

NOTICE OF EXEMPT APPROVAL

TO: Edna Jackson-Gray
Recruitment, School of Medicine WSU
1322 Scott Hall

FROM: Peter A. Lichtenberg, Ph.D. Peter A Lichtenberg
Chairman, Behavioral Institutional Review Board (B03)

DATE: December 17, 1999

RE: Protocol **#11-70-99(B03)-X**; "Effects of Coaching Methods Employed by Medical Students on Standardized Tests." No funding requested

The above named protocol and consent form have been reviewed and found to qualify for **Exemption** according to paragraph #2 of 45 CFR 46.101(b) of the Code of Federal Regulations of the Department of Health and Human Services.

As this proposal has not been evaluated for scientific merit, except to weigh the risk to the human subjects in relation to the potential benefits, this approval does not replace, or serve in place of, any departmental or other approvals that may be required.

NOTE:

Exempt Proposals do not require annual review by the IRB.

Please submit an Amendment Form to the HIC Office if there are any changes to the protocol while the study is being conducted.

Once the protocol has been completed, please submit a Closure Form to the HIC Office.

c: Faculty Supervisor, Department of Education, 397 EDU

APPENDIX B

Data Collection Instruments

SCHOOL OF MEDICINE

The information for this survey will assist programs to better serve students in preparing for standardized tests such as the USMLE.

Please complete this form and return it to: Recruitment Office – Room 1320

_____ _____ _____ _____ - ____ - _____
Last Name First Name MI Social Security Number

Please provide information regarding your preparation for and performance on the following standardized tests:

Have you taken the SAT or ACT :

1) I have taken the **SAT** ___yes ___no; I have taken the **ACT** ___yes ___no

2) If you have taken either or both tests, what were your scores?
If only "total" composite scores can be remembered, PLEASE give those.

SAT	Verbal	Math	Comp				ACT	Eng.	Math	Read	NS/SC	Comp
1st time							1st time					
2nd time							2nd time					
3rd time							3rd time					

3) How did you prepare for these tests?
(Please check "yes" or "no" for the test preparation method(s) that applies to you for the SAT and / or ACT)

	SAT			ACT	
Preparation Method(s)	Yes	No		Yes	No
Arco's Guide					
Barron's Guide					
Princeton Review Course					
Organized study groups by High School					
Organized study groups outside of school					
Organized workshops by High School					
Organized workshops outside of school					
Test Sample Booklet					
Other (please list)					

4) **NO preparation was done for the :_____SAT _____ACT**

MCAT PREPARATION

MCAT WAS TAKEN: ____ old (before 1991) ____ new (starting 1991)

MCAT PREPARATION METHOD(S) WAS/ WERE BY THE FOLLOWING:

For each time you took the MCAT, please indicate how much of each component (lecture, study guide, practice exams) you completed where: 100% is All; 75% is Most; 50% is half; 25% is "a little"; and 0% is "none".; NA is not applicable.

Barron's Guide

	Study Guide					Practice Exams					NA
	100%	75%	50%	25%	0%	100%	75%	50%	25%	0%	
1st time											
2nd time											
3rd time											
4th time											

Betz / Flower's Gujde

	Study Guide					Practice Exams					NA
	100%	75%	50%	25%	0%	100%	75%	50%	25%	0%	
1st time											
2nd time											
3rd time											
4th time											

Kaplan Course

	Lecture					Study Guide					Practice Exams					NA
	100%	75%	50%	25%	0%	100%	75%	50%	25%	0%	100%	75%	50%	25%	0%	
1st time																
2nd time																
3rd time																
4th time																

Princeton Review Course

	Lecture					Study Guide					Practice Exams					NA
	100%	75%	50%	25%	0%	100%	75%	50%	25%	0%	100%	75%	50%	25%	0%	
1st time																
2nd time																
3rd time																
4th time																

Excell Course

	Lecture					Study Guide					Practice Exams					NA
	100%	75%	50%	25%	0%	100%	75%	50%	25%	0%	100%	75%	50%	25%	0%	
1st time																
2nd time																
3rd time																
4th time																

Other (please list)

	Lecture					Study Guide					Practice Exams					NA
	100%	75%	50%	25%	0%	100%	75%	50%	25%	0%	100%	75%	50%	25%	0%	
1st time																
2nd time																
3rd time																
4th time																

References

Academic medicine and the national board: A love-hate relationship? Editorial.(1990). Academic Medicine, 65, 380-381.

Anderson, N. D. (1990). The mismeasure of medical education. Academic Medicine, 65, 159-160.

Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. American Psychologist, 36, 1086-1093.

Bangert-Drowns, R. L., Kulik, J. A. & Kulik, C. C. (1983). Effects of coaching programs on achievement test performance. Review of Educational Research, 53(4), 571-585.

Bangert-Drowns, R.L., Kulik, J.A. & Kulik, C.C. (1984, Jan). Synthesis of research on the effects of coaching for aptitude and admissions tests. Educational Leadership, 80-82.

Bangert-Drowns, R.L. & Others. (1984). Highlights from research on the effects of coaching for tests. (ERIC Document Reproductive Service No. ED 238 910).

Batter, W.T., Thomson, W. A., Morrissey, C.T., Miller, L.M. & Smith, D.W. (1991). Baylor's program to attract minority students & others to science & medicine. Academic Medicine, 66, 305-311.

Beaver, W. (1996). Is it time to replace the SAT? Academic Medicine, 82, 37-39.

Becker, B.J. (1990). Coaching for the scholastic aptitude test: Further synthesis and appraisal. Review of Educational Research, 60, 373-417.

Bell, G. (1994). The test of testing: Making appropriate and ethical choices in

assessment. (ERIC Document Reproduction Service No. ED 379 337).

Berliner, D. & Cassanova, U. (1986). Should students be made test-wise? Instructor. February, 22-23.

Bernstein, P.T. Tipping, J., Bercovitz, K. & Skinner, H.A. (1995). Shifting students and faculty to a PBL curriculum attitudes changed and lessons learned. Academic Medicine, 70, 245-247.

Blackburn, J.C. (1990). No one, including admissions offices, fail the SAT. College and University, 66, 17-20.

Boldt, R. F. (1986). Generalization of SAT validity across colleges. (ERIC Document Reproduction Service No. ED 279 711).

Boring, J.R. & Nutter, D.O. (1984). Analytic thinking; educating students for the practice of modern medicine. Journal of Medical Education, 59, 875-880.

Bracey, G.W. (1990). Do admissions officers fail the SAT? College and University, 66(16), 18-19.

Brazarsky, B., Frideman, C.P., Arnold, L., Davis, W.K., Jonas, H.S., Littlefield, J.H., & Martini, C.J.M. (1993). A view of medical practice in 2020 & it's implication for medical school admission. Academic Medicine, 68, 31-34.

Camacho, Z. & Thomson, W.A. (1991). Minority student admissions. Academic Medicine, 66, 604-605.

Cariaga-Lo, L.D., Enarson, C.E., Crandall, S.J., Zaccaro, D.J. & Ricards, B.F. (1997). Cognitive & noncognitive predictors of academic difficulty & attrition. Academic Medicine, 72, 569-571.

Carris, J. (1995). Coaching kids for the SAT. The Executive Educator, 17(1), 28-

31.

Chenoweth, K. (1996, September). SAT, ACT scores increase: Higher scores attributed to more rigorous coursework. Black Issues in Higher Education, 6-8.

Ciardi, E. (1990). Reading comprehension and the SAT. Journal of Reading, 33, 558-559.

Cohen, J.J. & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Association.

Cohen, J.J. (1996, November). Finishing the bridge to diversity. Presented at the plenary session of the 107th annual meeting of the AAMC, San Francisco, 6-12, 1996.

Cohen, S. A. & Hyman, J. S. (1991). Can fantasies become facts? Educational Measurement: Issues and Practices, 10, 20-23.

Cole, N. (1984, Fall). Testing and the "crisis" in education. Educational Measurement: Issues and Practice, 4-8.

Coles, C.R. (1985). Difference between conventional and problem based curricula in their students' approaches to studying. Medical Education, 19, 308-309.

Dempster, F. N. (1991, April). Synthesis of research on reviews and tests. Educational Leadership, 71-75.

Dolly, J.P. & Vick, D. S. (1986). An attempt to identify predictors of test-wiseness. Psychological Reports, 58, 663-672.

Drowns, R.L., Kulik, J.A., & Kulik, C.C. (1983). Effects of coaching programs on achievement test performance. Review of Educational Research, 53 (4), 571-585.

Drowns, R.L., Kulik, J.A., & Kulik, C.C. (1984, Dec/Jan). Synthesis of research on the effects of coaching for aptitude and admissions test. Educational Leadership, 80-

82.

Dyer, H. S. (1987, February). The effects of coaching for scholastic aptitude. NASSP, 46-53.

Edger, M. (1994). Measurement and evaluation. Studies in Education Evaluation, 20, 169-174.

Edwards, J.C., Johnson E.K., & Molidor, J.B. (1990, March). The interview in the admission process. Academic Medicine, 167-177.

Elam, C. L. & Johnson, M. M. S. (1997). An analysis of admission committee voting patterns. Academic Medicine, 72(Suppl.), 72-75.

Emmett, D. M. (1993). A descriptive analysis of medical school application forms. Academic Medicine, 68, 564-569.

Erdman, J. B. (1984). Coaching courses and the MCAT. The New England Journal of Medicine, 310, 388-389.

Federal Trade Commission. (1979). Effects of coaching on standardized analysis of data. Federal Commission. Washington, D.C. Bureau of Consumer Protection.

Franklin, E., Fleetwood, J. & Gracely, E. (1996). Taking cultural diversity seriously. Academic Medicine, 71, 2-3.

Frierson, H. T. (1986). Enhancing minority college students' performance on educational tests. Journal of Negro Education, 55(1), 38-45.

Gentry, J. M. & Perry, J. S. (1993). Test-wiseness, memory, and academic performance in university students. (ERIC Document Reproduction Service No. ED 375 351).

Graff, A.S. (1993). The new SAT: The future of transition assessment.

Educational Libraries, 17(3), 71-75.

Haladyna, T. M., Nolen, S. B. & Haas, N. S. (1991, June-July). Raising standardized achievement test scores and the origins of test score pollution. Educational Researcher, 2-7.

Hall, J. L. & Kleine, P. F. (1990). Preparing students to take standardized tests: Have we gone too far? (ERIC Document Reproduction Service No. ED 334 249).

Hall, M.L. & Stock, M.T. (1995). Relationship between quantity of undergraduate science preparation and preclinical performance in medical school. Academic Medicine, 70, 230-235.

Harris, K. H. & Longstreet, W. S. (1990, July-August). Alternative testing and the national agenda for control. The Social Studies, 148-152.

Harvill, L. M. (1984). Measuring the test-wiseness of medical students. Paper presented at the annual meeting of the American Educational Research Association (68th, New Orleans, LA). (ERIC Document Reproduction Service No. ED 244 999).

Hojat, M., Veloski, J.J. & Zeleznik, C. (1985). Predictive validity of the MCAT for students with two sets of scores. Journal of Medical Education, 60, 911-918.

Hymel, G. M. & Guedry-Hymel, L. (1987, October). Promoting study skills and test-taking techniques. NASSP, 97-100.

Hymes, D. L., Chafin, A. E. & Gonder, P. (1991). The changing face of testing and assessment: Problems and solutions. AASA Critical Issues Report. (ERIC Document Reproduction Service No. ED 339 726).

Hymes, D.L. (1993). Making sense of testing and assessment. (ERIC Document Reproduction Service No. ED 363 643).

Johnson, S. T. & Wallace, M. B. (1989). Characteristics of SAT quantitative items showing improvement after coaching among black students from low-income families: An exploratory study. Journal of Educational Measurement, 26(2), 133-143.

Jolly, P. (1992). Academic achievement & acceptance rates of underrepresented-minority applicants to medical school. Academic Medicine, 67, 765-769.

Jones, R. (1982, March). From MCAT to M.D.: Predicting success in medical school. Paper presented at the annual meeting of the American Educational Research Association, 66th. New York, N. Y., 19-23.

Jones, R. (1983). The relationship between MCAT science scores and undergraduate science GRADE POINT AVERAGE. Journal of Medical Education, 58, 908-1001.

Jones, R. (1984). MCAT scores and student progress in medical school. Journal of Medical Education, 59, 527-531.

Jones, R. (1984). Validity of the MCAT in predicting performance in the first two years of medical school. Journal of Medical Education, 59, 455-464.

Jones, R. (1986). A comparison of the predictive validity of the MCAT for coached and uncoached students. Journal of Medical Education, 61, 335-338.

Jones, R. (1986). The effects of commercial coaching course on performance on the MCAT. Journal of Medical Education, 61, 273-284.

Jones, F. & Flowers, J.C. (1990). New York's statewide approach to increase the number of minority applicants to medical School. Academic Medicine, 65, 671-674.

Keppel, G. & Zedeck, S. (1989). Data analyses for research designs analysis of variance and multiple regression/correlations approaches. New York: W.H. Freeman and

Company.

Kerlinger, F. N. (1986). Foundations of behavioral research. (3rd ed.). Forth Worth: Harcourt & Brace College Publishers.

Koenig, J. A. & Leger, K. (1997). Test-taking behaviors and their impact on performance. Academic Medicine, 72(Suppl.), 100-102.

Kouzekanani, K. & Others. (1989). Analysis of test wiseness components: A quasi-experimental approach. (ERIC Document Reproduction Service No. ED 306 239).

Lemann, Nicholas. (1999). The big test: The secret of the American meritocracy. New York: Farrar, Straus and Giroux.

Linn, R.L. (1990). Admissions testing: Recommended uses, validity, differential prediction, & coaching. Applied Measurement in Education, 3(4), 297-318.

Lynch, B.K., & Wood, M.K. (1990). The relationship of minority students' MCAT scores and grade point averages to their acceptance into medical school. Academic Medicine, 65, 480-482.

McGaghie, W.C. (1990). Qualitative variables in medical school admission. Academic Medicine, 65, 145-149.

McManus, B.L. (1992). The revised SAT's and the ACT's---- Are they really different? (ERIC Document Reproduction Service No. ED 352 368).

Mehrens, W. A. & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? Educational Measurement: Issues and Practice, 8(1). 14-21.

Messick, S. & Jungeblut, A. (1981). Time and method in coaching for the SAT. Psychological Bulletin, 89(2), 191-216.

Messick, S. (1982). Issues of effectiveness and equity in the coaching controversy: Implications for education and testing practice. (ERIC Document Reproductive Service No. EJ 254 500.

Millman, J. & Bishop, C. (1965). An analysis of test-wiseness. Educational and Psychological Measurement, 25(3), 707-726.

Mitchell, K. J. (1987). Use of MCAT data in admissions. A guide for medical school admissions officers and faculty. Association of American Medical Colleges. (ERIC Document Reproduction Service No. ED 310 689).

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

Mitchell, K.J. & Hynes, R. (1990). Score reporting for the 1991 Medical College Admission Test. Academic Medicine, 65, 719-723.

Morse, D. T. (1994, November). The relative difficulty of selected test-wiseness skills among college students. Paper presented at, the annual meeting of the Mid-South Educational Research Association. Nashville, TN., 9-11.

Moss, G. (1995). The effects of coaching on the ACT scores of African-American students. Paper presented at the annual meeting of the American Education Research association, San Francisco. (ERIC Document Reproduction Service No. ED 399 265).

National Center for Fair & Open Testing (Fairtest). (1992). How standardized testing damages education. (ERIC Document Reproduction Service No. ED 352 375).

Nunqester, R.J., Dawson-Saunders, B., Kelley, P.R. & Volle, R.L. (1990). Score reporting on NBME examinations. Academic Medicine, 65, 723-729.

Ornstein, A. C. (1993). Coaching, testing, and college admission scores. NASSP,

77, 12-19.

Paszczyk, S. L. (1994). A comparative analysis of ACT scores and final GPAs of Chicago State University undergraduate students. (ERIC Document Reproduction Service No. ED 370 519).

Petersdorf, R.C. Turner, K.S., Nickens, H.W., & Ready, T. (1990). Minorities in medicine: Past, present, & future. Academic Medicine, 65, 663-670.

Petersdorf, R.G. (1992). Not a choice, an obligation. Academic Medicine, 67, 73-79.

Popham, W.J. (1993). Task-teaching versus test teaching. Educational Measurement: Issues and Practice, 10-11.

Popham, W.J. (1991). Appropriateness of teachers' test-preparation practices. Educational Measurement: Issues and Practice, 12-15.

Powers, D. E. (1986). Relations of test item characteristics to test preparation/test practice effects: A quantitative summary. Psychological Bulletin, 100(1), 67-77.

Powers, D. E. (1987). Who benefits most from preparing for a "coachable" admission test? Journal of Educational Measurement, 24(3), 247-262.

Powers, D. E. & Leung, S. W. (1995). Answering the new SAT reading comprehension questions without the passages. Journal of Educational Measurement, 32(2), 105-129.

Prell, J. M. & Prell, P. A. (1986). Improving test scores-teaching test-wiseness: A review if the literature. (ERIC Document Reproduction Service No. ED 289 900).

Rainey, A. (1996). Effect of American College Testing (ACT) coaching course on American College Testing scores. (ERIC Document Reproduction Service No. ED

397 069).

Ready, T., (1995, January) Project 3000 by 2000: A model program to promote Hispanic education. Excellence in health sciences. Presented to the White House Commission on Hispanic Excellence in Education.

Report of the Project Panel on Other Important Considerations: Physicians for the 21st Century. (1984). Equity of access to a medical career. Journal of Medical Education, 59, 23-27.

Rickel, J. & Ruffin, A. (1995). Gender-associated differences in matriculating and graduating medical students. Academic Medical, 70, 552-559.

Rodriguez, R. (1996). Test-driven admissions. Black Issues in Higher Education, 13, 7-9.

Samson, G. E. (1985). Effects of training in test-taking skills on achievement test performance: A quantitative synthesis. Journal of Educational Research, 78, 261-265.

Sarnacki, R. E. (1979). An examination of test-wiseness in the cognitive test domain. Review of Educational Research, 49(2), 252-279.

Scott, L.K., Scott, C.W., Palmisano, P.A., Cunningham, R.D., Cannon, N.J., & Brown, S. (1980). The effects of commercial coaching for the NBME part Journal of Medical Education, 55, 733-742.

Seaton T. (1992). The effectiveness of test preparation seminars on performance on standardized achievement tests. (ERIC Document Reproduction Service No. ED 356 233).

Sedlacek, W.E. & Prieto, D.O. (1990). Predicting minority student's success in medical school. Academic Medicine, 60, 161-166.

Self, D. J. (1990). Moral dilemmas in medical school admission. Academic Medicine, 65, 179-183.

Shepard, L. A, (1990). Inflated test score gains: Is the problem old norms or teaching the test. Educational Measurement: Issues and Practice, 15-22.

Shepard, L. A. (1991). Effects of high-stakes testing on instruction. (ERIC Document Reproduction Service No. ED 337 468).

Smith, M. L. (1991). Meaning of test preparation. American Educational Research Journal, 28(3), 521-542.

Smyth, F. (1989, Spring). Commercial coaching and SAT scores. The Journal of College Admission, 2-9.

Snedecore, P. (1989, Fall). Coaching: Does it pay - revisited. The Journal of College Admission, 15-18.

Spooner, C. Jr. (1990). Help for the gatekeepers: Comment and summation on the admission process. Academic Medicine, 65, 183-157.

Sturgeon, T. V. (1994, Winter). Creating a standardized, systematic, and testable rating framework for competitive admission. Journal of College Admission, 6-13.

Swanson, A.G. (1984). Medical education in the United States and Canada. Physicians for the 21st Century. Journal of Medical Education, 59, 35-42.

Tabachnick, B. & Fidell, L. S. (1983). Using Multivariate Statistics. New York: Hayes and Row.

Taylor, R.E., Hunt, J.C., & Temple, P.B. (1990). Recruiting black medical students: A decade of effort. Academic Medicine, 65, 279-288.

Teague, A. B. (1992). A program to help college bound students improve their

verbal Scholastic Aptitude Test scores. (ERIC Document Reproduction Service No. ED 349 576).

Tekian, A. (1997). A thematic review of the literature on underrepresented minorities & medical training, 1981-1995: Securing the foundation of the bridge to diversity. Academic Medicine, 72, 5140-5146.

Thomas, M. C. (1986). Effective methods for increasing scores on standardized tests. (ERIC Document Reproduction Service No. ED 272 811).

Thompson, C. H. (1993). Education, race and class: A new calculus for the 21st century. Journal of Negro Education, 62(2), 113-124.

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

Weingartner, R. M. (1980). Selecting for medical school. Journal of Medical Education, 55, 922-927.

Weiss, M., Lotan, I., Kedar, H. & Ben Shakha, G. (1988). Selecting candidates for a medical school: An evaluation of a selection model based on cognitive and personality predictors. Journal of Medical Education, 22, 492-497.

Wilmouth, D. (1991). Should the SAT be a factor in college admissions? (ERIC Document Reproduction Service No. ED 345 592).

Worthen, B.R. & Spandel, V. (1991, Feb). Putting the standardized test debate in perspective. Educational Leadership, 65-69.

Yens, D. P. & Stimmel, B. (1982). Science versus nonscience undergraduate studies for medical school: A study of nine classes. Journal of Medical Education, 57,

429-439.

Yoder, E.L., (1997). Effect of medical student learner variable on transfer of learning. Unpublished doctoral dissertation, Wayne State University, Detroit.

Zebala, J.A. & Jones, D.B. (1989). Medical school admissions; The insiders guide. Revised Edition, chapter three The MCAT pp. 36-37. Memphis, TN; Mustang Publishing Co.

Zeleznik, C., Hojat, M. & Veloski, J. (1983). Baccalaureate preparation for medical school: Does types of degree make a difference? Journal of Medical Education. 58, 26-33.

Zeleznik, C., Hojat, M. & Veloski, J.J. (1987). Predictive validity of the MCAT as a function of undergraduate institution. Journal of Medical Education, 62, 163-196.

ABSTRACT

THE EFFECTS OF TEST PREPARATION METHODS EMPLOYED BY MEDICAL STUDENTS ON STANDARDIZED ADMISSION TESTS

by

EDNA E. JACKSON-GRAY

December 2000

Advisor: Donald Marcotte, Ph.D.

Major: Evaluation and Research

Degree: Doctor of Philosophy

Purpose of the Study: The primary focus of this study was to identify effective test preparation methods and components used by medical students to enhance their test score performance on the Medical College Admission Test (MCAT). Standardized tests such as the Scholastic Assessment Test (SAT) and American College Test (ACT) for college entrance and the MCAT for admission to most medical schools have an important impact on the education system in this country. Determining effective preparation methods could be beneficial for students at all academic levels.

Methods: Data pertinent to this study were collected from 383 first year medical students at a Midwest school of medicine using a self-report questionnaire. The population constituted students from two first year classes. Students indicated their method(s) of test preparation for college admission (ACT and/or SAT) and admission to medical school, the MCAT. They also estimated the amount of test preparation material completed per test preparation activity. Selected demographic data were subsequently collected from student files in a confidential manner with appropriate permission. Preparation methods

were categorized into two major types as self-directed (study manuals) and formal course (commercial coaching), which included study guides, lectures, and practice tests as components. The study was organized to examine four aspects of test preparation: academic levels of test preparation (college/medical school) SAT/ACT and MCAT method; types of test preparation (self-directed or formal course); components of the methods types of test preparation (lecture, study guide, practice tests) estimates of the percentage of components completed per method type used.

Results: Major findings indicated that examinees who participated in any form of test preparation out performed examinees for the Biological Sciences (BS) section of the MCAT, but not for the Physical Sciences (PS) and Verbal Reasoning (VR). Findings did not support evidence of a significant difference between MCAT scores of the students who had test preparation (SAT/ACT) for college admission and for medical school admission (MCAT) compared to those who only had test preparation for medical school admission.

Conclusions: MCAT scores appear to be positively affected by test preparation methods and techniques. The BS section of the test is most affected, with PS marginally, but not significantly affected. The VR section does not appear to be affected by test preparation.

AUTOBIOGRAPHICAL STATEMENT

Edna E. Jackson-Gray

Edna E. Jackson-Gray is a research assistant in the Minority Recruitment Office at Wayne State University School of Medicine - Detroit, Michigan. She graduated form Talladega College with a BA in Biology and from the University of Michigan with a MS in Occupational Education. She is also a certified Medical Technologists, MT (ASCP), completing her MT internship at Jewish Hospital, Cincinnati, Ohio.

Ms. Jackson-Gray has many presentations and workshop facilitations addressing study skills and test taking technique for high school and college students to her credit. She is an active member of several community service organizations and a number of professional and social organizations.