82-7

AN EVALUATION OF THE MULTISTATE BAR EXAMINATION

A Report Prepared for the

National Conference of Bar Examiners

Prepared by

Stephen P. Klein
GANSK & Associates
Los Angeles, California

August 30, 1982

PREFACE

The National Conference of Bar Examiners (NCBE) is a nonprofit organization composed of present and former members of boards of bar examiners throughout the country. One of NCBE's main goals is to protect the public by helping to insure that persons admitted to the bar are adequately equipped from the standpoint of knowledge and ability to serve as lawyers. In keeping with this goal, NCBE developed a 200 question multiple choice test called the Multistate Bar Examination (MBE). The first version of this test was administered in 1972. Since that time, the MBE has been included in the bar examinations of almost every state in the country. Over 50,000 applicants to the bar now take the MBE each year.

Several studies have been conducted on the MBE prior to those described in this report. These studies have shown that:

- MBE scores correlate well with scores on state developed essay and multiple choice bar examinations, with scores on the Law School Admissions Test (LSAT), and with grades in law school (Carlson and Werts, 1976; Klein, 1979).

- The MBE does not increase the difference in performance level between Anglo and minority groups that is present in their LSAT scores, law school grades, and bar examination essay scores (Klein, 1979 and 1981a).

- Lengthening the MBE's time limits improves scores only slightly and does not improve the performance of minority or generally low scoring applicants more than it improves the scores of other applicants (Klein, 1981b).

- MBE scores correlate well with scores on experimental measures of an applicant's ability to perform certain law practice related tasks, such as conducting legal research (Klein, 1981c).

The studies described in this report were undertaken to provide information about important aspects of the MBE that had not been already investigated. These studies were funded by NCBE as part of its mission to conduct research related to the bar admissions process. To insure the independence and objectivity of these studies, NCBE required that they be carried out by a technical research team and panels of attorneys who had not drafted MBE questions or been involved in the MBE's development.

Two types of studies are presented in this report: expert panel evaluations of various important substantive characteristics of the MBE, and statistical analyses of MBE data. The expert panels evaluated: whether the MBE's questions were material to the practice of law, whether they presented realistic case situations, whether they were allocated properly across content areas, and whether answering them correctly required appropriate degrees of legal knowledge and the ability to apply that knowledge. The statistical studies focused on question length, the relative attractiveness of the choices, whether some of the MBE's subtests (or items) were especially difficult for minority groups, and the adequacy of the test's time limits.

SUMMARY

The Multistate Bar Examination (MBE) is administered by almost every jurisdiction in the United States. This 200 item multiple choice test is generally used in conjunction with state developed essay tests to determine whether applicants to the bar have some of the important skills and knowledge that are considered necessary (albeit not sufficient) to practice law.

The MBE is developed by the National Conference of Bar Examiners and is administered under the direction of each jurisdiction's board of bar examiners. The Educational Testing Service prints, distributes, and scores the test. A new version of the test is constructed for each of its twice yearly administrations. About 50,000 applicants per year take the MBE.

This report describes three sets of studies that were conducted in order to provide an independent evaluation of various important characteristics of the MBE. The general design for these studies was developed by a team of technical consultants with nationally recognized expertise in the field of testing.

In the first set of studies, a joint committee from the Section of Legal Education of the American Bar Association, the Association of American Law Schools, and the National Conference of Bar Examiners selected five generalists, i.e., attorneys with expertise in several areas of the law. None of the panelists selected by this committee had ever been involved in drafting MBE items. The generalists individually and jointly evaluated several characteristics of all the items on the February 1978 version of the MBE. The major findings with the generalist panel were:

- Most of the items were judged to be material to the practice of law. Overall, 4 percent were rated low, 45 percent were rated high, and 51 percent were rated very high in materiality.

- Most of the items were judged to be appropriate for a test of basic competency. None of the items were considered to be clearly inappropriate for such a test, 12 percent were judged to have questionable appropriateness, and the remaining 88 percent were rated as appropriate to highly appropriate.

- In general, the panelists were satisfied with the realism of the case situations presented in the items, the levels of knowledge required in different content areas, and the relative emphases on memorization and analytic skills.

- The generalists agreed almost perfectly with the test developers in the relative allocation of items to the MBE's six content areas and they did not recommend expanding or contracting the number of areas assessed.

In the second set of studies, a panel of three experts was formed for each of the MBE's six content areas, i.e., Constitutional Law, Contracts, Criminal Law, Evidence, Real Property, and Torts. The 18 panelists were selected by the same committee that selected the generalists.

The members of each specialist panel individually and jointly evaluated several characteristics of the items in their area. These items were drawn from the February 1978 and July 1979 versions of the MBE. The major findings of the specialists were as follows:

- o Panelists concurred with the test developers in deciding which choice should be keyed correct on almost every item even though the panelists did not have the opportunity to discuss the rationale for the scoring key with the test developers.

- o In general, the panelists and the test developers were in close agreement as to the percentage of items that should be allocated to each topic within a content area.

- o Most of the items were judged to be material to the practice of law. Overall, 12 percent were rated low, 18 percent were rated medium, and 70 percent were rated high in materiality. This pattern was generally consistent across content areas.

- o The panelists felt that the items required about the right amount of emphasis on legal knowledge and analytic skills.

The third set of studies involved statistical analyses of examination data in order to supplement research that had already been done on the MBE. The database for these studies consisted of the responses of California examinees to the February 1978 and July 1979 versions of the MBE. The results of these studies indicated that:

- o An item's correct choice was usually chosen by 65 percent of the examinees. The most attractive incorrect choice was selected by 22 percent of the examinees. The second and third most attractive incorrect choices were selected by 9 and 4 percent, respectively. While it is consistent with good testing practices for an item's distractors to differ in attractiveness, the observed differences in the relative proportions were somewhat greater than desired.

- o There was no systematic relationship between the length of an item, i.e., the number words in it, and any of the following characteristics: the percentage of applicants who answered it correctly, the degree to which performance on the item was indicative of performance on the other 199 items, and the specialists' assessment of the item's materiality.

- o In general, there was no systematic relationship between the number of other items with which a particular question shared a fact pattern and any of the three characteristics noted above.

- o The relative differences among racial groups on one item was essentially the same as the differences among these groups on all the other items on the examination. In short, there was no evidence that certain items (or subtests) were especially easy or difficult for a particular group.

All of the foregoing findings are strongly supportive of the structure, format, and content of the MBE relative to its use as part of a state licensing examination program.

ACKNOWLEDGMENTS

The Board of Managers of the National Conference of Bar Examiners (NCBE) provided the resources and other support that were necessary for carrying out the research described in this report. NCBE also established an advisory committee that helped in planning and coordinating the research activities. This committee consisted of Douglas D. Roche (Chair), John Germany, Armando M. Menocal, III, and Joe Covington. The author was particularly appreciative of this committee's ability to provide assistance without interfering with the independence of the research or the interpretation of results.

Roger Bolus of GANSK & Associates and the University of California served as project manager. In that capacity, he helped in planning the research; arranging and conducting the panelist meetings; developing evaluation materials, directions, and rating forms; distributing these materials to panelists; tallying the results; and conducting statistical analyses of the data.

Earl Medlinsky and John Winterbottom of the Educational Testing Service were among the first to suggest that the study be done. In addition, they offered many useful suggestions for the research activities and provided data for some of the statistical analyses that were conducted. The Committee of Bar Examiners of the State Bar of California also provided some of the data used in this research.

A team of technical consultants with expertise in testing participated in designing the research and data analysis procedures. The members of this team and their institutional affiliations are described on page 2.

John Eckler of NCBE, Talbot D'Alemberte of the American Bar Association, and Millard Ruud of the Association of American Law Schools participated in the nomination and selection of the attorneys that served on the generalist and specialist panels.

Each panelist listed on page 5 and in Appendix A donated several days of his/her time to the project, including attendance at out-of-town meetings. The author also appreciated the effort and good nature they exhibited in carrying out their assigned tasks.

Everyone who participated in the project reviewed the initial version of this report. Their comments and suggestions were most helpful in identifying areas that needed revision. And, while the success of the project owes mainly to the contributions of these individuals, the author takes responsibility for any errors that still remain.

CONTENTS

TABLES

Chapter 1

INTRODUCTION

The purpose of a bar examination is to protect the public by requiring that persons licensed to practice law possess the basic legal knowledge and skills necessary for such practice. These skills include the ability to analyze fact situations, to apply fundamental principles of law in this analysis, to reason logically, to organize ideas, and to present these ideas clearly and persuasively. While an essay test can assess these skills, its breadth of coverage on any given administration is limited to a few issues in each content area. Moreover, the inherent nature of essay tests requires a certain degree of subjective judgment in grading answers.

MULTISTATE BAR EXAMINATION

The foregoing limitations of the essay format led the National Conference of Bar Examiners (NCBE) to develop the Multistate Bar Examination (MBE). The MBE was designed to supplement a state's own essay test by providing a reliable and objective measure of the ability to apply fundamental legal principles in analyzing fact situations in several content areas.

NCBE is responsible for drafting MBE items and making policy decisions regarding the use of this test. The Educational Testing Service provides editorial assistance in the drafting process; prints, distributes, and scores the test; and conducts statistical analyses of the results. The MBE is administered by the states choosing to use it in accordance with strict NCBE/ETS guidelines.

The MBE consists of 200 multiple choice questions (or "items"). Each MBE item has four choices. An applicant's raw score is the number of questions answered correctly. There is no correction for guessing.

The six areas covered by the MBE and the number of items in each area are: Constitutional Law (30), Contracts (40), Criminal Law (30), Evidence (30), Real Property (30), and Torts (40). Some of the questions share a fact pattern (case example) while other items stand by themselves. Items are not grouped by area within the test.

The items within each area are drafted by a team of attorneys. Each team consists of law professors, members of state boards of bar examiners, and other practicing attorneys. The items written by a team are reviewed by state boards of bar examiners, law professors, and testing experts. The suggestions of these external reviewers are considered by the drafting team and revisions in the items are made prior to their inclusion in the MBE.

A preliminary scoring of the answer sheets from several jurisdictions is performed shortly after the test is administered. This is done in order to identify items that did not appear to function properly (e.g., applicants who generally did well on the total examination did not select the choice that was keyed correct). The drafting teams review the items that are flagged by this procedure. They then determine the nature of any changes that need to be made in the scoring key (e.g., keying two or more choices on a given item as correct). Usually less than five items per test have their scoring key changed by this procedure.

A new form of the test is constructed for each administration. About 20 percent of the items in a form have appeared in a previous, but still secure, version of the MBE. These repeated items are used to adjust an applicant's raw score on the other 80 percent of the items so that the applicant's total score is not affected by possible differences in average item difficulty across administrations of the test.

The MBE is administered in two test sessions. Each session has 100 items (drawn proportionately from the six areas). Applicants are given three hours to complete each session and a break between sessions for lunch.

The first version of the MBE was administered in 1972. Since that time, it has been given twice per year, once in February and again in July. In 1981, about 54,000 applicants to the bar in 46 states and the District of Columbia took the MBE. Appendix B contains a more complete description of the MBE.

PURPOSES

In 1979, the National Conference of Bar Examiners commissioned a study that would provide an independent and external assessment of the quality of the MBE. Specifically, the study was designed to: (1) determine whether the MBE was measuring some of the skills and knowledge that were necessary (albeit not sufficient) for the practice of law, i.e., were MBE scores a valid indicator of the degree to which examinees possessed these skills and knowledge? and (2) identify ways in which the test could be improved.

GENERAL APPROACH

A team of nationally recognized experts in the field of testing was formed to design the study. The members of this technical consultant team are listed below:

Prof. Robert M. Guion
Bowling Green State University
Chair, Board of Scientific Affairs and
President elect, Division of Evaluation and Measurement
American Psychological Association

Dr. Stephen P. Klein
Senior Research Scientist
The Rand Corporation

Prof. Robert L. Linn
University of Illinois
President, the National Council of Measurement in Education
President, Division of Evaluation and Measurement
American Psychological Association

Prof. Jane R. Mercer
University of California, Riverside

Dr. Esteban L. Olmedo
Administrative Officer, Minority Affairs Program
American Psychological Association

The technical team determined that were was no generally acceptable and quantifiable index of an attorney's ability to practice law. Thus, it would not be possible to compare MBE scores with an appropriate measure of "on-the-job" performance. This situation led the technical team to recommend the formation of a panel of experts for each area covered by the MBE and that each panel evaluate a large sample of MBE items in its area. It also was decided to form a panel of generalists who would evaluate the test as a whole.

The technical team unanimously rejected the idea of forming a separate panel to make subjective judgments regarding whether certain MBE items might be biased against different sex and racial/ethnic groups. The reasons for this decision were: (1) the items undergo an extensive review for possible bias before they are included in the examination; (2) subjective judgments regarding bias, even by experts, rarely correspond to differential group performance (e.g., the difference in percentage of Anglo and minority examinees answering correctly an item rated as biased is usually the same as the difference between these groups on the items that are not rated as biased); and (3) statistical studies of the MBE have shown that differences in performance level between groups remains essentially constant across all the items in the test. However, the technical team did recommend replicating these statistical studies with a recent version of the test by means of standard and newly developed techniques for assessing bias.

Chapter 2 of this report describes the procedures that were used by the generalists and the results of their evaluations. Chapter 3 provides corresponding information on the six specialist panels. The final chapter discusses the procedures and results of the statistical studies that were recommended by the technical consultants and the panelists. These statistical studies were designed to supplement the research that had already been done on the MBE.

SELECTION OF PANELISTS

A three member committee was formed to participate in the nomination and selection of attorneys for the generalist and specialist panels. The members of this committee and the organizations that appointed them were as follows: John Eckler, National Conference of Bar Examiners; Talbot D'Alemberte, Section on Legal Education of the American Bar Association; and Millard Ruud, Association of American Law Schools.

Each committee member could veto the appointment of any proposed panelist. The only restrictions placed on who could be selected were that a panelist could not have been involved in the drafting of MBE items and could not be opposed in principle to the use of multiple choice questions on a bar examination. The selection committee was encouraged to form panels that not only exhibited content expertise but also had geographic, sex, and racial/ethnic diversity.

Chapter 3

GENERALIST PANEL

OVERVIEW OF ACTIVITIES

The Generalist Panel portion of the study had three phases. Phase 1 consisted of the nominating committee described in Chapter 1 selecting panelists and obtaining a commitment from them to participate in the research activities. The five panelists were:

E. Robert Blaske, Battle Creek, Michigan. J.D., University of Michigan, 1969. General practice of law with Blaske & Blaske. Chairman, Michigan Board of Law Examiners.

Charles D. Breitel, New York, New York. LL.B., Columbia Law School, 1932. Chief District Attorney, New York. Professor, Columbia Law School. Justice, New York Supreme Court. Chief Judge, New York Court of Appeals. Counsel, Proskauer Rose Goetz & Mendelsohn.

Ronald E. Kennedy, Chicago, Illinois. J.D., Northwestern University, 1973. Professor of Law, Northwestern University. President, Board of Governers, Chicago Council of Lawyers. Member, American Bar Association House of Delegates. Member, Board of Directors, Minority Legal Educational Resources.

Richard H. Montgomery, Seymour, Indiana. LL.B., Harvard Law School, 1950. General practice of law, Montgomery, Elsner & Pardieck. Member, State Board of Law Examiners. Member, Board of Managers, Indiana State Bar Association.

Wallace A. Tashima, Los Angeles, California. LL.B., Harvard Law School, 1961. Deputy Attorney General for the State of California. Member of the firm of Morrison & Foerster. Chair, California Committee of Bar Examiners. Judge, United States District Court.

In Phase 2, each panelist was mailed a copy of all the items that appeared on the February 1978 version of the examination. The generalists were asked to evaluate each item with respect to its materiality to the practice of law and its appropriateness in a test of basic competency for such practice. These evaluations were returned by mail to the project director.

In Phase 3, the panelists met to discuss and resolve differences in their Phase 2 evaluations, discuss other characteristics of the test, and make recommendations regarding: the percentage allocation of items to content areas, directions for future research, and ways in which the test could be improved.

RESULTS

Materiality

In Phase 2, the generalists were asked to evaluate all the items on the February 1978 test in terms of the degree to which the knowledge and skills they measured were material to the practice of law. In other words, did the items deal with trivial versus important legal issues and abilities? The panelists were further instructed that their rating of one item should be independent of their assessment of another item; e.g., they were not to mark an item as "low" simply because it was less material than some other item. These ratings were made on a scale from 1=Very Low to 5=Very High. The independent ratings of the five panelists on each item were averaged. Table 1 contains the percentage of items in each rating category.

Table 1

PERCENTAGE OF ITEMS IN EACH CATEGORY OF MATERIALITY

Rating Category	Average Score Range	Percentage of all Items
Very Low	1.00 - 1.99	0
Low	2.00 - 2.99	4
High	3.00 - 3.99	45
Very High	4.00 - 5.00	51

The generalists panel determined that 96 percent of the items had a high to very high degree of materiality to the practice of law. These results are very consistent with those obtained with the specialist panels (see Chapter 3) and strongly support the appropriateness of the use of the MBE in determining whether an attorney should be licensed to practice law.

Appropriateness for Testing Basic Competency

State bar examinations are not designed to predict who will become a good attorney. Rather, their goal is to identify those applicants to the bar who are not competent to practice (within the considerable limits of any testing device to make such a determination). This important distinction led to the decision to ask the generalists in Phase 2 to evaluate each MBE item in terms of the degree to which it measured skills and knowledge that would be appropriate for a test of basic competency to practice law. In short, were the skills and/or knowledge measured by an item generally necessary (albeit not sufficient) for practicing law in the broad content area covered by the item? Each generalist independently rated each item on a scale from 1=Highly Inappropriate to 4=Highly Appropriate.

The ratings of the five panelists on each item were averaged. Table 2 contains the percentage of items in each rating category. The data in this table indicate that the generalist panel determined that almost all the items were appropriate for a test of basic competency to practice law (the generalists' average rating across all the items was 3.35 out of a possible 4.00).

Table 2

PERCENTAGE OF ITEMS IN EACH CATEGORY OF APPROPRIATENESS FOR TESTING BASIC COMPETENCY

Rating Category	Average Score Range	Percentage of all Items
Inappropriate	1.00 - 1.99	0
Questionable	2.00 - 2.99	12
Appropriate	3.00 - 4.00	88

Other Dimensions

Each generalist was asked to review all 200 items on the February 1978 test, identify items he thought presented especially realistic case situations, and identify items he felt presented the most unrealistic situations. The panelists then met and discussed the characteristics of both types of items. Following this discussion, each panelist independently evaluated the test as a whole in terms of the general level of realism of the case situations.

The generalists repeated the foregoing steps, i.e., review, identification of illustrative items, discussion, and independent ratings, for three other dimensions: the level of knowledge required to answer the items, the degree to which the items emphasized memorization of legal principles, and the degree to which the items emphasized analytic skills.

The ratings of all four dimensions were on a five point scale from 1=Very Low to 5=Very High. Table 3 presents the panelists' average score on each dimension. These data indicate that the generalists felt that the items generally had moderate to high degrees of realism and that they required moderate to high degrees of knowledge of general legal principles and analytic skill. They further felt that the items required only a low to moderate degree of memorization of specific information.

Table 3

OVERALL RATINGS OF ITEM CHARACTERISTICS

Characteristic Evaluated	Average Rating
Realism of Case Examples	3.7
Level of Knowledge	3.5
Memorization	2.8
Analytic Skills	3.8

The generalists' discussion of the February test led them to make the following recommendations and comments with respect to the characteristics listed in Table 3:

- o Items should generally strive to be realistic, i.e., involve fact situations that arise frequently in daily practice or at least could arise.

- o Some items should require "spot knowledge" of key principles and rules.

- o No effort should be made to change the general level or range of legal knowledge required by the test. The items should not be made easier or more difficult. However, some of the more simple items, those requiring no more than common sense to answer, should be eliminated.

- o It was appropriate for certain areas, such as Real Property, to require the examinee to possess a higher degree of legal knowledge than other areas because of the varying degree to which specific knowledge was required for practice in each area.

- o The general level and balance of memorization and analytic skills required by the test were about right.

The foregoing results are generally consistent with those in Table 6 in that both the generalists and the specialists determined that hardly any of the MBE's items required overly advanced skills or knowledge.

Allocation of Items to Content Areas

The generalists were asked to give their independent assessment of the percentage of items that should be allocated to each of the MBE's six content areas. These allocations and the percentage of items in each area that actually appear on the test are presented in Table 4. An inspection of these data indicates an extremely close correspondence among the panelists. There also is a high degree of consistency between their average allocation to an area and the percentage of items in that area on the examination.

The generalists also were asked to make recommendations regarding whether the MBE should be expanded to include any additional areas. This request led to a discussion of the importance and appropriateness of several areas that were not now included in the examination (none of the panelists recommended dropping any of the existing areas). Some of the areas considered were: Civil Procedure, Taxation, Administrative Law, Family Law, Remedies, Agency, Trusts, and Corporations. Of these, only Civil Procedure was nominated by at least two panelists. However, it was agreed that jurisdictional differences in these procedures would probably preclude adding a Civil Procedure section to a nationally used test. Family Law faced the same problem. It was further agreed that if taxation were going to be tested, then it should involve an open book type of examination. Thus, the consensus of the group was that the current allocation of MBE items to areas should not be changed.

Table 4

PERCENTAGE ALLOCATION OF MBE ITEMS TO SUBTESTS

Subtest	Panelist Number 1	2	3	4	5	Average Across All Panelists	Actual Allocation
Constitutional Law	15	20	15	10	20	16	15
Contracts	20	15	20	20	20	19	20
Criminal Law	15	20	15	10	20	16	15
Evidence	15	15	15	20	10	15	15
Real Property	15	15	15	20	10	15	15
Torts	20	15	20	20	20	19	20

Chapter 3

SPECIALIST PANELS

OVERVIEW OF ACTIVITIES

The Specialist Panel portion of the study had three phases. Phase 1 involved selecting three panelists for each of the MBE's six content areas and obtaining a commitment from them to participate in all the research activities. Appendix A lists the panelists in each area.

In Phase 2, each of the 18 panelists was mailed a set of 30 to 40 items in his/her assigned content area. These items were drawn from the February 1978 version of the MBE (which was chosen because it was the most recent publicly released form of the test). The panelists were asked to evaluate each item with respect to several characteristics (e.g., its materiality to the practice of law) and then return these evaluations by mail to the project director.

In Phase 3, each of the six teams of panelists met to discuss and resolve differences in their Phase 2 evaluations, evaluate items from the July 1979 version of the MBE, and make recommendations regarding the ways in which the test could be improved.

RESULTS

Correctness of Scoring Key

In Phase 2 of the activities described above, panelists were asked to identify the choice they considered to be the best answer to each item on the February 1978 test. At least two of the three members of each panel agreed with the keyed response on 86 percent of the items.

The agreement rate rose to 95 percent in Phase 3 when the panelists had an opportunity to discuss with one another why they felt certain answers were correct. It is very likely that the level of agreement with the scoring key would have increased still further had the panelists been able to discuss the rationale for the keying with those who drafted the items.

Most of the disagreements with the scoring key occurred on the Criminal Law and Real Property subtests (see Table 5). An analysis of the 9 items on which agreement with the scoring key was not reached indicated these items tended to be more difficult than the other items on the test in terms of the percentage of applicants who answered them correctly. However, applicants who chose the keyed choice on an item where disagreements occurred had a higher average total test score than applicants who did not select the correct choice (the average biserial correlation on the 9 items was essentially the same as the overall average biserial correlation).

Two other factors should be considered in interpreting these findings: (1) all MBE items undergo extensive review prior to their inclusion in the test as well as statistical analysis after their administration but prior to assigning scores in order to detect any errors in the scoring key (see Appendix B), and (2) the overall level of agreement among raters was not especially high on any of the dimensions evaluated; i.e., the reliability problem was not restricted to judging correctness (see Appendix C).

Table 5

PERCENTAGE OF ITEMS ON WHICH PANELISTS AGREED WITH THE SCORING KEY BEFORE AND AFTER DISCUSSING WHICH CHOICES SHOULD BE KEYED CORRECT

Content Area	Independent Evaluations	After Group Discussion
Constitutional Law	80	97
Contracts	80	100
Criminal Law	87	87
Evidence	100	100
Real Property	80	87
Torts	88	100
Average	86	95

Allocation of Items to Topics

The items within each of the six subtests can be drawn from several topic areas. For example, items in the Contracts subtest can deal with consideration, conditions, remedies, etc. In Phases 2 and 3, the panelists were asked to review the list of topics in their subtest area that was used by the drafting team, add any additional topics they thought should be included, and then indicate the percentage of items in the subtest that should be allocated to each topic. They were further instructed to make these allocations on the basis of their assessment of the relative importance of the topics to the practice of law.

Table 6 presents the average difference in the percentage of items a panel recommended allocating to a particular topic and the percentage of items that were actually devoted to that topic on February 1978 examination. An inspection of these data indicates that there was generally a fairly close correspondence between the panelists' and the test developers' allocation of items to topic areas. For example, on any given topic, the two groups usually differed by only 7 percent (about 2 items on a 30 item subtest).

Table 6

DIFFERENCES BETWEEN PANELISTS AND TEST DEVELOPERS IN THE PERCENTAGE OF ITEMS THEY WOULD ALLOCATE TO A GIVEN TOPIC

Content Area	Number of Topics	Average Difference	Largest Difference
Constitutional Law	5	10.0	15.5
Contracts	10	5.2	13.0
Criminal Law	6	3.5	9.0
Evidence	6	7.8	22.5
Real Property	6	5.0	10.0
Torts	5	9.2	17.0
Average	6	6.8	14.5

The largest disagreements between the panelists and the test developers would lead the panelists to recommend changing the current allocations in the following ways:

- o Constitutional Law - add more items to Individual Rights.
- o Contracts - add more items to Parol Evidence and Interpretation and subtract items from Contract Formation and Consideration.
- o Criminal Law - add more items to Constitutional Protection and subtract items from Inchoate Crimes.
- o Evidence - add more items to Presentation of Evidence and subtract items from Relevancy.
- o Real Property - include items dealing with Land Finance and Land Use Control.
- o Torts - add more items to Defamation and Privacy, and subtract items from Negligence.

Some of the differences between the test developers' and the panelists' allocations may have stemmed from the need to include in a given version of the examination items from previous administration of the test (in order to provide a basis for scaling scores). The relatively short length of each subtest and the use of multiple items for a single fact pattern further constrained the number of different topics that could be covered in a single administration. In other words, there is likely to be some variation across test forms in the percentage allocation of items. And, if a different set of panelists were selected, they would likely disagree with at least some of the allocations made by the panelists in this study. Thus, while there was relatively close agreement between the panelists and the test developers on the allocation of items to most topics, it was not at all surprising to observe some differences between them. What is important, however, is that the size of these differences were generally quite small.

Materiality

The panelists were asked to evaluate the items in their assigned area in terms of the degree to which the knowledge and skills they measured were material to the practice of law (from 1=very low to 5=very high). In other words, did the items deal with trivial versus important legal issues and abilities? Each item on both the February 1978 and July 1979 examinations was evaluated on this dimension. Table 7 contains the percentage of items on each subtest on each examination that the panelists placed in the "medium to high" category of materiality. Overall, about 12 percent were rated low, 18 percent medium, and 70 percent as high in materiality. The low percentage for Contracts on the February examination in Table 7 was due mainly to one panelist. This panelist apparently reversed the rating scale (so that 24 of the 40 items were rated as low or very low in materiality) because this panelist gave just the opposite ratings of materialiy to the July items. Thus, the overall percentage of items rated as medium to high is probably over 90 percent.

Table 7

PERCENTAGE OF ITEMS RATED MEDIUM TO HIGH IN MATERIALITY

Content Area	February	July	Average
Constitutional Law	92	93	93
Contracts	55	87	71
Criminal Law	76	86	81
Evidence	93	100	96
Real Property	93	80	87
Torts	97	97	97
Average	84	91	88

Legal Knowledge and the Ability to Apply it

The panelists also were asked in Phase 3 to evaluate the July 1979 items in terms of (1) the level of legal knowledge (from basic to advanced) that is required in order to answer the questions correctly and (2) the extent to which an examinee had to use reasoning skills in applying legal knowledge to the facts presented. A three point scale from 1=low to 3=high was used for making both types of evaluations.

On both dimensions, about 25 percent of the items were placed in the low category, 68 percent in the medium category, and less than 10 percent in the high category. Table 8 presents the percentage of items in each area that were rated in the low category on each dimension.

It is not clear whether the massing of items in the medium category was due to the panelists' true assessment of the items, the ability of the directions to explain the rating scales, or simply the tendency on the part of the panelists to avoid extreme rating categories. In any event, it is evident that the panelists felt strongly that the items did not generally require unusually high or low degrees of knowledge and/or reasoning skills.

Table 8

PERCENTAGE OF ITEMS RATED LOW IN REQUIRING LEGAL KNOWLEDGE AND IN APPLYING LEGAL KNOWLEDGE IN ORDER TO BE ANSWERED CORRECTLY

Content Area	Legal Knowledge	Ability To Apply
Constitutional Law	13	20
Contracts	35	20
Criminal Law	43	30
Evidence	24	10
Real Property	23	30
Torts	18	26
Average	26	23

Correlations of Rating Scales with Item Statistics

An analysis of the data indicated that there was only a weak inverse relationship between the percentage of examinees answering a question correctly and the degree to which the panelists felt that item required specialized legal knowledge and (to a lesser extent) advanced analytic skills. The reverse was true for materiality (i.e., easier items tended to have slightly higher materiality ratings than more difficult items). This pattern of relationships was most apparent on the Contracts, Evidence, and Real Property subtests.

There was no systematic relationship between how well an item performed (as indicated by how well it predicted an applicant's total score) and and the ratings assigned to that item by the panelists.

Panelist Comments

After completing all the activities described above, the panelists were asked to provide written recommendations regarding any aspect of the examination. Most of the comments that were made were directed toward their assigned subtest and included suggestions for additional topic areas. The general recommendations made by two or more panels are summarized below:

- Increase the realism of the fact situations; especially by drawing upon the kinds of cases an examinee is likely to encounter in actual practice. For instance, one member of the criminal law panel felt that the need to use "jurisdictionless" items led to unrealistic case examples. This panelist suggested providing examinees with a particular statutory pattern and then asking them several questions that are based on this pattern.

- Increase the consistency of the grammatical form in which the questions are posed to applicants.

- Use descriptive terms like "Buyer" and "Seller" instead of names like Bob and Sam.

In general, the panelists thought that the existing approach and coverage of the examination was good, but that there was room for improvement. They also suggested some revisions in the organization of topics within an area (most of which have already been reflected in the 1981 examinations).

Chapter 4

STATISTICAL ANALYSES OF EXAMINATION DATA

The technical consultants and the panelists suggested that certain studies be conducted to supplement research that had already been done on the MBE. They recommended investigations of: the relative attractiveness of the choices on each item, the effect of item length on examinee performance, the effect of the number of items in a fact pattern on test statistics, the extent to which differences among racial groups remained constant across subtests and items, and the appropriateness of the test's time limits.

The database for the studies described in this chapter consisted of the responses made by the examinees who took the February 1978 and/or July 1979 administration of the MBE in California. This database was used because of the large number of applicants who took the test in this jurisdiction, the availability of item and race data on these examinee, and the use of these tests with the specialist and generalist panels.

ATTRACTIVENESS OF ITEM CHOICES

The Item Choice Study investigated the following issues: (1) did examinees generally pick the correct choice, and (2) what was the relative pulling power of the "distractors" (choices that were keyed incorrect).

Table 9 presents the percentage of examinees selecting the correct choice as well as those choosing the first, second, and third most attractive distractor on all the items in the February 1978 and July 1979 versions of the examination. These data indicate that the correct choice was usually chosen by about 65 percent of the applicants and that the distractors were not equally attractive. This general pattern was present on all subtests.

Analyses indicated that applicants who selected the correct choice on an item had a higher average total score on the other items than those who selected the most attractive distractor. And, those who selected the most attractive distractor had a higher average score than those that selected the second or third most attractive distractor. These findings indicate the more able examinees were eliminating incorrect choices that less able examinees found attractive. Thus, applicants who knew more were better able to eliminate clearly incorrect choices.

Table 9

AVERAGE PERCENTAGE OF EXAMINEES SELECTING EACH TYPE OF CHOICE

Type of Choice	February	July	Average
Correct Choice	64	66	65
Most Attractive Distractor	23	21	22
2nd Most Attractive Distractor	9	9	9
3rd Most Attractive Distractor	4	4	4

EFFECT OF ITEM LENGTH

There are three major ways in which the length of MBE items differ from one another:

- o The number of words in the fact pattern on which the item is based.
- o The number of words in the item (including the number of words in each alternative).
- o The total number of words that have to be read per item, including the item's proportionate share of the fact pattern on which it is based.

These three characteristics of an item were correlated with its difficulty (the percentage of examinees who answered it correctly), its discriminability (the extent to which examinees who answered it correctly also tended to answer the other items correctly), and its materiality (as judged by the specialist panels). The results of this analysis indicated that there was generally no systematic relationship between any of the dimensions of an item's length and its difficulty, discriminability, or materiality. The only exceptions to this trend were that longer Criminal Law and Torts questions tended to have slightly better discriminability and higher ratings of materiality than shorter questions. There were no statistically significant negative correlations; e.g., there was no evidence that longer items tended to be more difficult or have lower discriminability and materiality than shorter items.

EFFECT OF THE NUMBER OF ITEMS IN A FACT PATTERN

In general, there was no statistically significant correlation between an item's characteristics (its difficulty, discriminability, and materiality) and the number of other items with which it shared a fact pattern. The two exceptions to this trend were: Criminal Law items tended to have higher discriminability if they shared a fact pattern with other items and Torts items tended to have higher materiality if they shared a fact pattern with other items. It also was observed that performance on items that shared a fact pattern tended to correlate only very slightly higher with performance on the other item(s) in that fact pattern than with performance on the other items in the test.

ITEM BIAS

Past research on the MBE has indicated that the differences in performance level between sex and racial groups in one area, such as Torts, are generally consistent with the differences between these groups in other areas (Klein, 1976). In other words, the questions in one area are not <u>especially</u> difficult for a particular group (relative to how well that group generally performs on items in other areas). However, since these studies were conducted several years ago, it was decided to replicate them with a current version of the examination.

The first step in conducting this analyses involved drawing a random 10 percent sample of Anglo applicants who took the July 1979 version of the MBE (N = 561). This sample was then combined with all the Asian (N = 256), Black (N = 376), and Hispanic (N = 403) applicants who took this test. California applicants were used for this purpose because California is one of the few states that collects racial data on its applicants and it has the largest number of applicants in each racial group.

The second step involved conducting an analysis of the degree to which differences in average performance among groups paralleled differences among them on each of the other five subtests. The statistical technique used for this purpose was a repeated measures analysis of variance in which Race (4 levels) and Subtest (6 levels) were the independent variables.

The results of this analysis indicated that far less than one percent of the differences in scores among applicants was due to one subtest being particularly difficult for one or more groups; i.e., the relative differences among groups stayed almost perfectly constant across the six subtests even though the subtests themselves varied in difficulty (see Table 10). For instance, about 70 percent of the Anglo applicants but only 64 percent of the minority applicants answered a typical Criminal Law question correctly; i.e., a difference of six percent. However, there also was about a six percent difference between Anglo and minority applicants on each of the other five subtests. And, the relative standings of the groups was consistent across all six subtests.

Table 10

PERCENTAGE OF ITEMS ON EACH SUBTEST THAT WERE ANSWERED CORRECTLY BY THE APPLICANTS IN EACH RACIAL GROUP

	Const Law	Crim Law	Con-tracts	Evi-dence	Real Prop	Torts	Average
Anglo	68	70	71	70	61	67	68
Asian	62	65	67	67	56	64	63
Hispanic	61	64	66	65	55	62	62
Black	60	63	65	64	54	60	61
Anglo vs. Minority Average	7	6	5	5	6	5	6

Mean Subtest Score = % answered correctly X number of items

A set of six repeated measures analyses of variance were run for the items within each subtest using the procedures described by Cleary and Hilton (1968). In each analysis, Item (30 to 40 levels depending upon the subtest) and Race (4 levels) were the independent variables. The score earned by each applicant on each item (where 0 = wrong and 1 = correct) was the dependent variable. The results of this analysis were identical to those described above; i.e., none of the items within a subtest were especially difficult or easy for a particular group.

The six analyses of variance described above considered each item within a subtest relative to the other items in that subtest. Recent developments in the field of testing have provided methods for assessing whether a particular item is biased relative to a group's performance on the total examination. This type of analysis involves determining whether examinees of comparable ability (but belonging to different groups) had the same success on any given item. In this context, comparable ability is defined as having similar total scores on the examination. The procedures used to conduct this analysis with each of the 200 items on the July 1979 examination are summarized below (see Ironson & Subkoviak, 1979; and Scheuneman, 1979 for a more technical discussion of these procedures):

1) Examinees in each of the four racial groups used in the analyses of variance described above were classified into five ability groups based on the overall distribution of total MBE scores.

2) The number of examinees in each racial group at each ability level was tabulated.

3) The percentage of examinees in each ability group (regardless of race) who answered a given item correctly was computed.

4) The data from steps 2 and 3 were used to calculate the percentage of applicants in each combination of race and ability group that would be expected to answer the item correctly.

5) The actual percentages from step 2 were compared to the expected percentages from step 4 by means of a chi square test.

The results of the foregoing analyses indicated that none of the items had a statistically significant chi square value ($p = .05$); i.e., there was not a single item on which performance differences between racial groups systematically deviated from the general pattern of performance differences between these groups. This finding corroborates the analysis of variance results described previously which indicated that there was absolutely no evidence of bias due to race. It must be noted, however, that all of the foregoing analyses were run on California applicants. No data were available to assess whether the results obtained would hold up with other minority groups, such as Hispanic applicants of Puerto Rican descent.

TIME LIMITS

Analyses of MBE data have found that almost all applicants answer every question in the time allotted (e.g., Dorans and Wright, 1981). This could occur because essentially all applicants had enough time to attempt each item. It also could occur because most applicants realize they can maximize their score by using the last minute of the morning and afternoon test sessions to mark randomly one choice for each item they had not yet attempted to answer (because there is no correction for guessing). In other words, if applicants did not have enough time to complete the MBE, there would be far more guessing on the questions towards the end of each test session than there would be towards the beginning or middle of a session.

If guessing occurred more often towards the end of a test session, then: (1) scores on the items appearing at the end of a test booklet would be lower than those appearing near the beginning and middle of the booklet, and (2) scores on the items at the end of a test booklet would not be as highly correlated with total scores as would scores on the items that appeared towards the beginning and middle of the booklet (because they would be more affected by chance).

Tables 11 and 12 contain data on item difficulty (deltas) and correlations with total scores (biserials) for the February and July tests given in 1981. These data indicate that there was no consistent tendency for the items towards the end of a test booklet to be more difficult or have lower correlations with total scores than items towards the beginning and middle of a test booklet. Thus, there is no statistical evidence that guessing occurs more often towards the end than towards the beginning or middle of a test session which in turn suggests that the time limits are adequate for completing the examination. These findings are consistent with an experiment (Klein, 1981b) that found that there would be only a very slight improvement in raw scores (about 3 points per 100 items) if applicants had essentially unlimited time to answer MBE questions.

Table 11

CHARACTERISTICS OF ITEMS ANSWERED AT THE BEGINNING, MIDDLE, AND END OF EACH TEST SESSION OF EACH EXAMINATION GIVEN IN 1981

Test Session	Item Number	Mean Equated Delta		Mean r-biserial	
		Febr	July	Febr	July
Morning	1-33	10.7	10.6	.26	.30
	34-66	11.3	11.2	.25	.32
	67-100	11.5	11.5	.26	.32
Afternoon	101-133	11.1	11.5	.32	.33
	134-166	12.4	11.4	.27	.29
	167-200	11.7	11.3	.23	.29
Average	1-200	11.5	11.3	.27	.31

Mean Equated Delta = index of item difficulty (the higher the delta, the more difficult the item). The average standard deviation of the deltas within a block of 33 items was 1.89 for February and 2.23 for July. There was no systematic relationship between the standard deviations and the sequence of blocks.

Mean r-biserial = index of item discriminability (the higher the r-biserial, the stronger the relationship between an applicant's score on the item and that applicant's total score on the other 199 items in the test). The average standard deviation of the r-biserials within a block of 33 items was .11 for both February and July. There was no systematic relationship between the standard deviations and the sequence of blocks.

Table 12

CORRELATION OF ITEM STATISTICS WITH SERIAL POSITION

Test Session	Correlation with Mean Equated Delta		Correlation with Mean r-biserial	
	Febr	July	Febr	July
Morning	-.15	.06	.31	-.05
Afternoon	-.10	.11	.10	-.08
Average	-.13	.08	.20	-.06

The data in this table indicate that the higher the item number on the morning session of the February examination, the more likely it was to have a higher biserial. None of the other correlations with an item's serial position were statistically significant.

APPENDIX A

MEMBERS OF THE SPECIALIST PANELS

Constitutional Law

Justice Arthur J. England Jr., Tallahassee, Florida
Dean James C. Kirby Jr., University of Tennessee
Francis P. Mood, Columbia, South Carolina

Contracts

Carole Kamin Bellows, Chicago, Illinois
Stuart Duhl, Chicago, Illinois
Prof. Curtis R. Reitz, Philadelphia, Pennsylvania

Criminal Law

*E. Robert Blaske, Battle Creek, Michigan
Prof. Robert J. Levy, Minneapolis, Minnesota
Mary P. Walbran, Owatonna, Minnesota

Evidence

C. A. Powell III, Birmingham, Alabama
Maurice Reuler, Denver, Colorado
Prof. M. Michael Sharlot, Austin, Texas

Real Property

Jerome Hafter, Greenville, Mississippi
Henry Kordes, Sea Girt, New Jersey
Prof. Edward H. Rabin, Davis, California

Torts

Prof. Carl S. Hawkins, Provo, Utah
Martin L. Gross, Concord, New Hampshire
Justice Robert P. Smith Jr., Tallahassee, Florida

* Also served on the Generalist Panel

Appendix B

DESCRIPTION OF THE MBE

GENERAL CHARACTERISTICS

The Multistate Bar Examination (MBE) is a 200 item (question) multiple choice test. Each item contains four choices. The six areas covered by the MBE and the number of items in each area are: Constitutional Law (30), Contracts (40), Criminal Law (30), Evidence (30), Real Property (30), and Torts (40). The items in an area are distributed throughout the test rather than presented together. As many as 4 items may deal with the same fact pattern (reading passage) although many items stand by themselves.

The MBE is divided into two sessions. The 100 items given in each session are drawn proportionately from the six areas. Applicants are given 3 hours per session. An applicant's raw score is the total number of questions answered correctly; there is no correction for guessing. Almost all applicants answer all items.

The MBE is administered twice per year, once in February and again in July. About 160 new items are constructed for each administration. The remaining items are drawn proportionately from the six areas from a previously administered but still secure version of the MBE. The items from a past administration are used for equating, i.e., transforming raw scores to scale scores. The transformation process adjusts the raw scores for possible differences in average item difficulty across administrations. Thus, a given scale score denotes a level of proficiency that is not affected by the relative difficulty of the particular set of items that were answered.

There is no national passing score on the MBE. Instead, each state sets its own policies regarding standards for passing and how MBE raw or scale scores are combined with scores on the state developed portion(s) of the bar examination.

TEST DRAFTING AND REVIEW

There are six MBE drafting teams, one per area. Each team consists of two bar examiners and three law professors with expertise in their assigned content area plus a test development specialist. One of the law professors serves as the team's chairperson. The NCBE's director of testing recruits team members and monitors their activities.

A team's first activity involves preparing specifications for its area. The specifications describe the major topics and concepts to be covered. The topics and concepts chosen are ones that are commonly included in a basic course in the area, are important for legal practice, and can be measured adequately by a multiple choice test. The specifications are reviewed and, if necessary, revised annually.

The next step involves the team members individually drafting items. Written guidelines for the item drafters emphasize clarity, conciseness, and fairness. There is never an intent to trap or trick applicants into the wrong answer.

The first draft of each item is sent to ETS where it is reviewed and edited to insure consistency with the MBE's style and format. The complete set of revised items is sent to all team members prior to their next meeting.

At this meeting, the team members review and revise items, designate those that are rejected, those that are preliminarily accepted, and those that require further revision before they can again be considered for inclusion. Items that eventually appear in the MBE undergo two such reviews. An assessment is then made of the difficulty of the preliminarily accepted items. In addition, the team reviews alternative sets of items from past administrations in order determine which set should be used for equating. Sets of equators are chosen on the basis of content and statistical considerations. Some time also is spent on reviewing the characteristics of items they drafted for previous administrations. This is done to help team members estimate item difficulties and improve their item writing skills.

A draft test is prepared for each area. The items in this draft are selected by the team (or team chairperson) on the basis that they conform to the specifications for content coverage and span a range of item difficulties. The items then undergo an editorial review. The reviewer's questions are resolved by the test consultant or the team (if the questions are substantive in nature).

The six drafts are assembled into AM and PM test books. The construction of these books takes into consideration content coverage and arrangement, estimated item difficulties, the placement of equating items, and other important factors. For instance, there is balance with respect to the frequency with which each choice (A, B, C, or D) is keyed correct and, to avoid confusion, a name that appears in one item is not used again in another unrelated item. The items also are checked and, if necessary, revised to make sure they avoid any language or situational description that is likely to be offensive to subgroups of applicants.

A mock up version of the test is constructed and sent to the teams for review. Scoring keys are not provided during the initial phase of this review. Instead, team members answer the questions in their area as if they were applicants taking the test for the first time. They then discuss the items and suggest additional revisions if they feel problems still exist.

The revised mock ups are sent to interested state boards of bar examiners and to NCBE's MBE committee. Comments by these external reviewers are forwarded to the appropriate teams. The teams review the comments and decide whether further revisions are warranted and, if so, what they should be. The final version of the test is then constructed and the answer key and other documentation are once again carefully checked to make certain that all changes have been made correctly.

The final version of the test is printed and distributed to the states in accordance with policies and procedures that help to ensure test security. The states administer the MBE under guidelines designed to maintain security, standardization, and fairness.

EARLY ITEM ANALYSIS

A sample of answer sheets is analyzed prior to processing all of them. This "early" item analysis involves checking that the computer programs used for scoring the answer sheets are consistent with the scoring key. It also involves flagging items that are answered incorrectly by applicants who otherwise perform well on the test. Each flagged item is reviewed by ETS, NCBE, and the appropriate team chairperson. As a result of this review, the answer key for a flagged item could be changed or more than one of its choices could be keyed correct. This final safeguard on test quality and fairness usually results in modifying the scoring key for about five items.

SCORING AND SCORE REPORTING

The revised scoring key is used for grading all the answers, i.e., including the ones used in the early item analysis. The scoring is done by machine, but at least 0.5% of the answer sheets are hand graded to monitor accuracy. There also are several other quality control checks in the procedures used for transforming raw scores to scale scores and for preparing the final roster of applicant identification numbers and grades. Only after all of these checks have made is an applicant's score sent to the state in which he/she took the examination.

The following scores can be obtained: total raw score (the number of questions answered correctly), raw score in each of the six areas, and total scale score. States are strongly encouraged by both NCBE and ETS to use just the total scale score in making pass/fail decisions.

APPENDIX C: STATISTICAL TABLES

Table C.1

AVERAGE INTERRATER CORRELATIONS

	Generalists		Specialists		
Area	Basic Competency	Materiality	Materiality	Knowledge	Reasoning
Constitutional Law	.25	.30	.07	.19	.48
Contracts	.49	.53	.36	.62	.39
Criminal Law	.35	.47	.37	.23	.64
Evidence	.21	.00	.59	.24	.24
Real Property	.14	.26	.61	.15	.33
Torts	.35	.38	.41	.30	.32
Average	.30	.32	.40	.29	.40

The generalists rated each February 1978 item in terms of its appropriateness for a test of basic competency and the degree to which the knowledge and skills it measured were material to the practice of law. The specialists rated each July 1979 item on this latter dimension as well as in terms of the level of legal knowledge that is required to answer it correctly and the extent to which an examinee had to use reasoning skills in applying legal knowledge to the facts presented.

Table C.2

CORRELATIONS OF SPECIALISTS' RATINGS WITH ITEM CHARACTERISTICS

	Materiality		Knowledge		Reasoning	
Area	Percent Correct	r-bis	Percent Correct	r-bis	Percent Correct	r-bis
Constitutional Law	.17	.02	-.29	-.09	-.27	-.05
Contracts	.28	.15	-.54	-.23	-.40	-.31
Criminal Law	.14	.01	-.18	-.02	.15	-.27
Evidence	.17	.18	-.49	-.03	-.37	.05
Real Property	.33	-.15	-.46	-.12	-.22	.01
Torts	.28	-.02	.06	-.21	.21	-.15
Average	.23	.03	-.32	-.12	-.15	-.12

The average of the three panelists' ratings of an item on a dimension was correlated with the percent answering that item correctly and that item's correlation with the sum of the scores on the other items on the test (i.e., its r-biserial correlation). Results indicated a tendency for easy items to be judged more material to the practice of law and require less reasoning skills than difficult items.

Table C.3

FREQUENCY DISTRIBUTION OF CHI SQUARE VALUES IN ITEM BIAS ANALYSIS

Chi Square	Frequency	Chi Square	Frequency
0 - 1.0	11	9.1 - 10.0	11
1.1 - 2.0	14	10.0 - 11.0	13
2.1 - 3.0	26	11.1 - 12.0	10
3.1 - 4.0	20	12.1 - 13.0	5
4.1 - 5.0	12	13.1 - 14.0	3
5.1 - 6.0	22	14.1 - 15.0	3
6.1 - 7.0	16	15.1 - 17.0	4
7.1 - 8.0	17	17.1 - 20.0	4
8.1 - 9.0	9	20.1 - 21.0	1

Chi square critical (at alpha = .05) = 21.03

REFERENCES

Carlson, A. B. and Werts, C. E. Relationship among law school predictors, law school performance, and bar examination results. Princeton, N.J.: Educational Testing Service, 1976.

Cleary, T. A. and Hilton, T. L. An investigation of item bias. Educational and Psychological Measurement, 1968, 28, 61-75.

Dorans, N. and Wright, R. Test Analysis: Multistate Bar Examination, Forms 3CEB2 and S-3CEB2. Unpublished Statistical Report, SR-80-117. Princeton, N.J.: Educational Testing Service, October, 1980.

Ironson, G. H. and Subkoviak, M. J. A comparison of several methods of assessing item bias. Journal of Educational Measurement, 1979, 16, 209-225.

Klein, S. An investigation of possible item and grader bias in a state bar examination. Paper presented at the meetings of the Western Psychological Association, Los Angeles, April, 1976.

Klein, S. An analysis of the relationships between bar examination scores and an applicant's law school, admissions test scores, grades, sex, and racial/ethnic group. Paper presented at the American Bar Association meetings, Dallas, Texas; August 14, 1979. Reprinted in The Bar Examiner, 1980, 49, 14-18.

Klein, S. Factors associated with the difference in passing rate between Anglo and Hispanic applicants on the New Mexico Bar Examination. Report prepared for the New Mexico Board of Bar Examiners, 1981a.

Klein, S. The effect of time limits, item sequence, and question format on applicant performance on the California Bar Examination. Report prepared for the Committee of Bar Examiners of the State Bar of California and for the National Conference of Bar Examiners, 1981b.

Klein, S. Testing research skills on the California Bar Examination. Report prepared for the Committee of Bar Examiners of the State Bar of California and for the National Conference of Bar Examiners, 1981c.

Scheuneman, J. A method of assessing bias in test items. Journal of Educational Measurement, 1979, 16, 209-225.