TESTING RESEARCH SKILLS ON THE CALIFORNIA BAR EXAMINATION

A Report Submitted to the
Committee of Bar Examiners of the State Bar of California
and the National Conference of Bar Examiners

Prepared by
Stephen P. Klein
GANSK & ASSOCIATES

December 31, 1981
PREFACE

In 1979, the Committee of Bar Examiners (CBE) of the State Bar of California held a series of conferences with experts in legal education and testing. These meetings focused on identifying the kinds of skills that should and could be measured by a bar examination, especially those that were important in actual legal practice. It was apparent from these discussions that many of the skills and new assessment techniques that were considered warranted further investigation.

The National Conference of Bar Examiners (NCBE) concurred with this view and with their financial assistance, the CBE undertook the development and field testing of several new measures of lawyering skills (e.g., the ability to interview clients and the ability to represent them competently in legal proceedings). These tests were given in conjunction with the July 1980 administration of the California bar examination.

This report, which deals with the Research Test, is the first of a series of reports describing the results with the new measures. As its name implies, the Research Test assesses some of the skills that are required for carrying out legal research (such as the ability to determine whether existing case law can be used to support a client’s case). One unique feature of this test is that examinees are given copies of the cases and statutes on which they are to base their answers; i.e., it is analogous to an open book examination.

Several other NCBE/CBE supported studies also were conducted in conjunction with the July 1980 administration of the California examination. These studies focused on a variety of issues associated with the current form of the examination, such as the degree to which scores on the multiple choice and essay portions of it are affected by the amount of time applicants are given to complete these sections. The results of these studies are presented in another report (Klein, 1981a).
SUMMARY

The July 1980 version of the California Bar Examination was divided into two parts. One part, called the General Bar Examination (GBX), had two sections; the Multistate Bar Examination (MBE) and the Essay Examination. The MBE had 200 multiple choice questions and all the applicants participating in the research described in this report had MBE scores. The Essay section had nine questions, however, because of California's multiphased grading process, about 30 percent of the applicants had only three of their essay answers graded.

The other part of the July 1980 examination was called the Special Session (SS). The SS was optional; i.e., an applicant could choose not to take it without jeopardizing his/her chances of passing the GBX. If an applicant participated and did well in the SS, it could increase that applicant's chances of passing. About 99 percent of the applicants who sat for the complete GBX also took the SS.

Applicants were assigned randomly to one of the SS's four sections. Since these sections were administered concurrently, a given applicant could take only one of them. This report describes one of the SS's four sections, namely the Research Test, and presents the results obtained with this test.

The Research Test was designed to measure an applicant's legal research skills. Three forms of the test were developed, however, a given applicant took only one of them. Each form had two parts, A and B. Part A required the applicant to evaluate on a five point scale the degree to which each of several cases supported various legal propositions that were relevant to a client's case. Part B asked the applicant to use the information from Part A and other materials, such as relevant and irrelevant statutes, in order to prepare a memorandum about several facets of a client's case, such as its strengths and weaknesses. Since applicants were given copies of all the cases to be evaluated, the test emphasized analytic skills rather than legal knowledge. Each applicant was given 195 minutes to complete the test.

Results indicated that the ratings the applicants made in Part A and the scores that were assigned to the memoranda they wrote in Part B provided reliable indices of their performance levels. The correlation between the Part A and B scores was about the same as that obtained between two questions on the Essay section of the GBX. The Part A scores correlated slightly higher with the scores on the MBE than with scores on the Essay portion of the GBX while the reverse was true for the Part B scores. Total scores on the Research Test correlated with total GBX scores only slightly less well than the MBE and Essay scores correlated with each other.

The Research Test did not widen or narrow the gap in performance levels between racial groups that is present on the GBX. Similarly, repeater status, type of law school from which the applicant graduated, and most other background characteristics were just as highly correlated with Research Test scores as they were with scores on the GBX. For example, applicants who had worked as law clerks and received pay for doing legal research prior to taking the Research Test tended to have higher scores on this test (and on the GBX) than applicants who did not have this training. These findings do not necessarily mean that such experience caused the higher test scores in that the applicants who are selected for these jobs also may be the ones with higher than average law school grades.
Questionnaire data indicated that applicants thought the Essay and MBE portions of the GBX were better measures of their legal knowledge than was the Research Test. However, they thought the Research Test was a better measure of their ability to perform as an attorney. They also felt that the case situations presented in the Research Test were far more realistic than those used in either the Essay or MBE sections of the GBX. The applicants' chief complaint about the Research Test was that they did not have enough time to complete it. There was no relationship between an applicant's assessment of the quality of the Research Test, MBE, or Essay sections of the examination and that applicant's scores on these sections.

The pattern of relationships with scores on one form of the Research Test was essentially the same as the pattern of relationships with scores on the other two forms of the test. These findings along with those described above suggest that the Research Test measures important skills that are similar but not identical to those measured by the MBE and Essay portions of the GBX. They also suggest that a revised version of the Research Test could be a valuable addition to the GBX, especially if its development cost could be reduced and spread across several jurisdictions.
ACKNOWLEDGMENTS

The Committee of Bar Examiners of the State Bar of California provided the advice and cooperation that were necessary for carrying out the studies described in this report. Two members of this committee, Armando M. Menocal, III and Martin R. Glick, made especially important contributions in the design, implementation, analysis, and report preparation phases of this research. John O’Hara provided valuable advice and assistance throughout the project. He was also instrumental in making it a jointly supported effort of the Committee of Bar Examiners and the National Conference of Bar Examiners.

Professor Frederick M. Hart of the University of New Mexico School of Law and Professor Leo O’Brien of Hastings Law School were primarily responsible for developing the test materials used in this research.

The Committee's staff reviewed, pilot tested, edited, printed, and administered the test materials. This staff, along with the Committee’s Board of Reappraisers, also selected, trained, and calibrated the cadre of lawyers who graded the essay portions of the tests. Kenneth D. McCloskey, the Committee's former Director of Testing, was in charge of and participated actively in all of these activities. The efficiency with which these tasks were completed, the quality of the data obtained, and the fact the research activities were carried out concurrently with testing and processing over 7,500 applicants on the regular bar examination testify to the skills, energies, and unstinting dedication that Ken provided. Ken was ably assisted in these activities by James B. Tippin, Jr., Suzanne M. Obermeier, and Philip Schoner. Phil also provided invaluable help in coordinating readers, data cleaning, and records management.

John Bianchini and Andrew York of the Educational Testing Service developed and implemented the computer systems that were used in assigning applicants to groups and in cleaning and linking diverse data files. Randy Onishi and Roger Bolus of GANSK & Associates also participated in the computer data cleaning and file management activities. Roger was further responsible for conducting the statistical analyses presented in this report.

Ralph Hoepfner of the System Development Corporation provided a thorough and extremely helpful technical review of a draft of this report. He also participated in the development of one of the test forms used in the study.

Finally, the project could not have been carried out without the excellent cooperation of the applicants who participated in it. Their tolerance and good nature in putting up with the many unique demands of the research at a time when they were under great emotional strain were most appreciated by all who were involved in this project.
CONTENTS

PREFACE ................................................................. iii

SUMMARY ............................................................. v

ACKNOWLEDGMENTS ....................................................... vii

Chapter
  1. INTRODUCTION .................................................. 1
     Background ..................................................... 1
     Purpose ......................................................... 2

  2. GENERAL PROCEDURES .......................................... 3
     Overview ....................................................... 3
     General Bar Examination ...................................... 3
     Special Session ............................................... 5
     Background Characteristics ................................... 6

  3. DESCRIPTION OF THE RESEARCH TEST ........................... 8
     Overview ....................................................... 8
     Development History ......................................... 8
     Format and Instructions ....................................... 8
     Materials ....................................................... 9
     Time Limits .................................................. 9
     Scoring ......................................................... 10

  4. PRELIMINARY ANALYSES ......................................... 11
     Comparability of Samples .................................... 11
     Item Analysis of Part A ....................................... 11
     Reader Consistency in Part B ................................ 12
     Reliability of Total Score ................................... 15

  5. RESULTS .......................................................... 16
     Relationship Between Research and GBX Scores ............... 16
     Relationship of Research Test Scores to Other Measures ... 16
     Attitudes Towards the Test ................................... 18

  6. SUMMARY AND IMPLICATIONS OF FINDINGS ....................... 19

REFERENCES .......................................................... 21

Appendix
  A. Excerpts from Notice to Applicants Regarding Special Session 22
  B. Pre-examination Questionnaire ................................ 23
  C. Post-examination Questionnaire ............................... 25
Chapter I

INTRODUCTION

BACKGROUND

The major goal of a bar examination is to identify those applicants to the bar who have the skills and knowledge that are necessary for practicing law. In a recent article in The Bar Examiner, O'Hara and Klein (1981) indicated that some of the skills that are frequently mentioned as being necessary are the abilities to:

- Analyze legal problems
- Collect and sort facts
- Marshall facts
- Understand and interpret opinions and regulations
- Write effectively
- Conduct legal research

Many of these skills are related to one another and some, such as the ability to analyze legal problems, are already measured well by the multiple choice and essay portions of most state bar examinations. However, current bar examinations do not measure some of the other important skills listed above, such as the ability to do legal research.

Critics of bar examinations also claim that the existing tests place too much emphasis on an applicant's ability to memorize legal principles and rules because in actual practice, an attorney often consults reference materials to obtain this information. Both the multiple choice and essay portions of the examination further require that the applicant deal with several different case situations in a relatively short period of time. This is an effective testing strategy because an applicant's score is not unduly influenced by knowledge (or lack of knowledge) about a particular legal principle. In other words, it allows for a wide sampling of case situations. However, one consequence of this strategy is that applicants are not tested on their ability to handle a substantial segment of a client's case.

The foregoing considerations led to the development of a test that would measure several important legal research skills. This test was designed to assess an applicant's ability to: assess the degree to which a wide array of case law and statutes supported a client's position, marshal the facts in support of that client's case, identify strengths and weaknesses in the case, and determine additional information that would be needed and the probable sources of that information. An important feature of this test was that it was not heavily dependent on specific legal knowledge. Thus, an applicant taking this test was provided with all of the case law to be analyzed, copies of statutes, etc.
PURPOSE

The analyses described in this report were conducted to answer the questions below:

- Are scores on the Research Test related to scores on the essay and multiple choice sections of the General Bar Examination?
- Are scores on the Research Test related to an applicant's legal training and experience?
- Are scores on the Research Test related to an applicant's attitudes about this test and/or the GBX?
- Is the difference in average performance level between Anglo and minority applicants on the Research Test larger or smaller than it is between these groups on the General Bar Examination?
- Do applicants think the Research Test is a realistic and good measure of their legal skills?
Chapter 2

GENERAL PROCEDURES

OVERVIEW

The July 1980 version of the California Bar Examination was divided into two parts. One part, called the General Bar Examination (GBX), had two sections; the Multistate Bar Examination (MBE) and the Essay examination. Most of the applicants who took one section also took the other section.

The other part of the July 1980 examination was called the Special Session (SS). The SS was optional; i.e., applicants could choose not to take it without jeopardizing their chances of passing the GBX. However, if an applicant participated and did well in the SS, it could increase that applicant's chances of passing. About 99 percent of the applicants who sat for the GBX also took the SS.

The SS had four sections that were administered concurrently. A given applicant could, therefore, take only one of these sections. The focus of this report is on the section involving the Research Test. Information about the background characteristics of the applicants who took this test was obtained from state bar records and questionnaires.

GENERAL BAR EXAMINATION

The July 1980 General Bar Examination (GBX) was administered over a three day period. The first day and the afternoon of the second day were devoted to the Essay. The MBE was given on the third day.

Multistate Bar Examination

The Multistate Bar Examination (MBE) is developed by the National Conference of Bar Examiners and is published and scored by the Educational Testing Service. About 45 states and the District of Columbia use the MBE as part of their bar examinations.

The MBE consists of 200 multiple choice questions (or "items"). The six areas covered by the MBE area are as follows: Constitutional Law, Contracts, Criminal Law, Evidence, Real Property, and Torts. Some of the questions relate to a common set of facts while other items stand by themselves. Items are not grouped by area within 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.

Each MBE item has four choices. An applicant's raw score is the number of questions answered correctly; i.e., there is no correction for guessing. The raw scores are converted to "scale scores" by the Educational Testing Service so that an applicant's final score is not affected by possible differences in overall test difficulty across administrations of the test. California multiplies these scale scores by 3.0 so that the theoretical maximum score on the MBE in California is 600 points.
Essay Examination

The Essay portion of the July 1980 GBX contained nine questions. These questions were developed by law professors teaching at non-California schools and by the California Committee of Bar Examiners and its staff. Some of the questions focused on only one subject matter area while others, called "cross-over questions," covered two areas.

The Essay section was divided into three test sessions of three hours each with three questions per session. Applicants were allowed to allocate the three hours in any way they wished across questions within a session.

The grading of the essay answers was done by a team of 12 lawyers per question (i.e., there were a total of 108 readers). These readers were trained to use the same criteria and standards in evaluating the answers to their question. The answers were graded in five point increments on a scale of 0 to 100 points. A score of 70 was assigned to an answer if the reader was in doubt as to whether the answer was a minimally adequate response to the question.

Pass/Fail Decisions

There are two ways an applicant can pass the GBX. Method 1 consists of receiving a combined Total score (i.e., MBE + Essay) of 1050 or greater. Method 2 involves receiving an MBE score of 420 or greater on one administration of the GBX and an Essay score of 630 or greater on another administration; i.e., in order to pass by the second method, an applicant must pass both sections of the examination. If an applicant fails the GBX because the applicant's score on one section was not high enough to compensate for a failing score on the other section, the applicant can retake the entire GBX (and thereby have the opportunity to pass under either method) without jeopardizing the passing status on the section passed previously.

Multiphased Grading

Research on California bar examinations (Klein, 1978 and 1980a) has indicated that final pass/fail decisions can be predicted with 99.5 percent accuracy without scoring all the essay answers written by many applicants. This finding and other considerations led the Committee of Bar Examiners to adopt a multiphased grading process. In Phase I of this process, 3 of an applicant's 9 essay answers are graded. The scores on these answers are combined with the applicant's MBE score to produce a Phase I score. This score is used to separate applicants into two groups, namely: pass and questionable. In subsequent phases, applicants in the latter category have their remaining 6 essay answers graded (and when necessary, regraded) in order to make accurate pass/fail decisions.

One consequence of the multiphased system is that total Essay scores were available for only 70 percent of the applicants taking this section. The group with complete data were not representative of the total applicant pool. Thus, for the purposes of this research, a total Essay score was created for all applicants by multiplying by 3.0 their total score over the three questions that they had graded in Phase I of the multiphased system. If an applicant did not have at least three essay questions graded, then no Essay score was created for that applicant.
Summary Statistical Data

Table 2.1 presents summary statistical data on the July 1980 GBX for the 7379 applicants who took both the MBE and Essay portions of this test. The reliability of the MBE was computed by Dorans and Wright (1980). The reliability of the Essay portion was computed by using the Spearman-Brown formula to step up the average inter-question correlation within sessions \( r = 0.32 \). It was not possible to compute the average correlation across all questions because the multiphased grading process resulted in some of the applicants not having all of their essay answers read and those who did have all of them read were not representative of all the applicants who took the examination. The reliability of the total GBX score was estimated using the procedures described by Gulliksen (1950) for a linear composite. There was a .88 correlation between the MBE and Essay sections.

Table 2.1

<table>
<thead>
<tr>
<th>Statistic</th>
<th>MBE</th>
<th>Essay</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Score</td>
<td>424.5</td>
<td>617.5</td>
<td>1042.0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>46.0</td>
<td>54.4</td>
<td>92.2</td>
</tr>
<tr>
<td>Reliability</td>
<td>.88</td>
<td>.81</td>
<td>.91</td>
</tr>
</tbody>
</table>

Some of the other characteristics of the applicant pool were: 49 percent passed the GBX, 69 percent took the GBX for the first time, and 62 percent were graduates of American Bar Association approved law schools. Within this latter group, 81 percent took the GBX for the first time. The passing rate among these ABA first timers was 73 percent. These results are very consistent with those obtained with past administrations of the examination (Klein, 1981b).

SPECIAL SESSION (SS)

Prior to the administration of the examination, applicants were advised that there would be a Special Session (SS). They were further advised that (1) their scores in the SS would not be counted if they passed the regular examination (i.e., based on the total of their Essay and MBE scores) and (2) their scores in the SS would be counted as one-sixth of their total grade if they failed the regular examination (see Appendix A). Thus, participating in the SS could increase but not decrease their chances of passing. Applicants also were informed that they had to take the MBE and Essay portions of the GBX in order to derive any benefit from taking the SS and that they would be assigned randomly to one of its four sections. In other words, a given applicant was assigned to one of four concurrently administered sections.

The SS was administered on the morning of the second day of examination. Of the 7379 applicants who took the complete GBX, 98.5 percent elected to participate in the SS.
A stratified random sampling plan was used to assign applicants to sections. The stratification variables included race, type of law school from which an applicant graduated, and repeater status. The number of applicants assigned to a section was based on how many would be needed for statistical analyses of the data and the administration and scoring costs of each section. This report focuses on one of the four sections of the SS, namely, the Research Test. This test is described in Chapter 3.

BACKGROUND CHARACTERISTICS

Law School Type and Repeater Status

All persons applying to take California's bar examination are required to provide a transcript of their law school grades. These transcripts made it possible to determine the type of law school from which an applicant graduated; i.e., American Bar Association (ABA) approved, California Accredited (but not ABA approved), and Unaccredited. State bar records also were used to determine whether an applicant had taken and failed the GBX previously; i.e., whether or not the applicant was a repeater.

Sex and Racial/Ethnic (R/E) Group

All applicants taking the bar examination in California are requested to complete a form on which they indicate their sex and racial/ethnic (R/E) group affiliations. This information, which is used solely for research purposes, was provided by 99.6 percent of those taking the July 1980 examination. The distribution of sex and racial/ethnic groups on this examination was as follows: males (71%) and females (29%); Anglos (82%), Asians (4%), Blacks (6%), and Hispanics (6%). The Asian group is composed mainly of Chinese and Japanese Americans, but it also includes those with Philippino and Pacific Islander backgrounds. The Hispanic group is largely Mexican-American, although it contains small percentages of applicants with Central American, South American, and Puerto Rican backgrounds.

Questionnaire Data

Applicants were advised that in order to derive any benefit from their participation in the Special Session (i.e., in terms of passing the GBX), they had to complete and return the questionnaire that appears in Appendix B. This questionnaire requested information about certain potentially relevant background characteristics, such as the extent to which English was spoken in the home, and data about an applicant’s legal training and experience.

All applicants who applied to take the July 1980 examination were mailed a copy of the questionnaire prior to the examination. Applicants who did not return the questionnaire by early July were sent another copy of it. They also were advised again that it had to be completed in order for them to derive any benefit from the Special Session and that they could turn it in at the examination site. Of the 7379 applicants taking the complete GBX (and thereby eligible to benefit from the Special Session), 96 percent returned questionnaires. And, almost all of those who completed both the full GBX and the Special Session returned questionnaires.
A questionnaire also was administered between the morning and afternoon sessions on the third day of the examination. This questionnaire inquired about the applicants' preparation for the examination; assessment of the adequacy of the time limits for the MBE, Essay, and Special Session; and opinions about how well each of these sections measured their legal skills and knowledge. A copy of this questionnaire appears in Appendix C. About 95 percent of the applicants who participated in the SS and took the complete GBX turned in this questionnaire.
Chapter 3

DESCRIPTION OF THE RESEARCH TEST

OVERVIEW

An applicant taking the Research Test section of the Special Session was assigned randomly to one of the three forms of this test. Each form had two parts, A and B. Part A required the applicant to evaluate the degree to which various cases supported different legal propositions. Part B asked the applicant to use the information from Part A and other materials to prepare a memorandum about several facets of a client's case.

DEVELOPMENT HISTORY

The idea for the Research Test grew out of discussions among the members of the Committee of Bar Examiners of the State Bar of California and several consultants to the Committee. The first form of the test, which dealt with a case involving medical malpractice, was developed and field tested in the fall of 1979 by Professor Frederick Hart at the School of Law, University of New Mexico. This field testing led to several revisions in the test's format, instructions, length, and content.

The revised version of Form 1 served as the model for constructing Form 2. This form, which dealt with a case involving consumer fraud, was developed by Professor Leo O'Brien of Hastings Law School.

Form 2 and the revised version of Form 1 were pilot tested in May, 1980 with attorneys who had passed the California examination. These attorneys had varying degrees of experience in conducting legal research. The results of the May testing led the Committee of Bar Examiners and its staff and consultants to make a few more revisions in both forms of the test. Form 3 of the test, which was a slight variant of the first form, also was constructed at the time these changes were made.

FORMAT AND INSTRUCTIONS

Each test form had two parts, designated as A and B. In Part A, the applicant was given several cases and a set of propositions. Each proposition, such as "An accused must be allowed to impeach prosecution witnesses by means of juvenile adjudications," was discussed in at least one of the cases. Applicants were instructed to evaluate the degree to which a case supported a proposition using the following five point scale.

1=Holding and express rationale contradict the proposition
2=Holding OR express rationale (but not both) contradict the proposition
3=Holding and express rationale are NEUTRAL with respect to the proposition
4=Holding OR express rationale (but not both) support the proposition
5=Holding and express rationale support the proposition
Part A was identical on Forms 1 and 3. It consisted of all possible combinations of 5 cases and 9 propositions; i.e., a total of 45 case/proposition combinations. Part A on Form 2 had 40 case/proposition combinations. However, since Form 2 dealt with a fact pattern involving consumer fraud while Forms 1 and 3 were related to issues involving medical malpractice, Form 2's cases and propositions were entirely different from those used with Forms 1 and 3.

The instructions for Part B on all three forms placed the applicant in the role of an associate in a law firm. These instructions, which were in the form of a memorandum from a senior partner in this firm, asked the applicant to draft memoranda in response to questions about a particular client's case. Applicants were further instructed to base their memoranda on the cases they analyzed in Part A, plus additional materials (including factual information they were given about the client's case).

Part B of Form 1 asked for five separate memoranda in response to specific questions. Some of these questions appear below:

- What information available from the materials indicates that Dr. Brown may have been negligent?
- What additional factual information should our firm attempt to obtain in order to prove liability in negligence? What are the likely sources of this information?
- Assuming that we are successful in establishing liability in negligence, what damages are recoverable?

Part B of Form 3 embedded all of Form 1's questions within a memorandum from the senior partner to the applicant (this was the only difference in the content and instructions of these two forms).

Part B of Form 2 was like Part B of Form 3 in that it contained several questions embedded within a memorandum from the senior partner to the applicant.

MATERIALS

An applicant taking any one of the Research Test's three forms received the following materials for that form: General Directions, Directions for Part A, List of propositions for Part A, a booklet containing cases for Part A, a form for recording answers to Part A, Directions for Part B, Additional materials for Part B, and a booklet in which to write the memoranda for Part B.

TIME LIMITS

Applicants were given all of their test materials at the beginning of the session. They were then told they had 195 minutes to complete the session. They also were told they could allocate 195 minutes any way they wished across parts, but that they should probably devote somewhat more than 75 minutes to Part B.
Part A

The test developers along with the staff and consultants of the Committee of Bar Examiners developed a scoring key for each version of Part A. This key indicated the most appropriate rating for each case/proposition combination. A score was assigned to each rating using the values in Table 2.2 so that the closer the applicant's rating came to the keyed response, the higher score the applicant received for that rating. For example, if an applicant rated a given case/comination as a 1 and the correct choice was a 2, the applicant would receive 4 points for that case/proposition combination. A rating of 3 was rarely keyed correct.

Table 2.2

PART A ITEM SCORES FOR EACH COMBINATION OF APPLICANT RATING AND KEYED RESPONSE

| Applicant's Rating | Keyed Response |
|--------------------|----------------|----------------|
| 1                  | 5  4  3  2  1  |
| 2                  | 4  5  4  3  2  |
| 3                  | 3  4  5  4  3  |
| 4                  | 2  3  4  5  4  |
| 5                  | 1  2  3  4  5  |

The sum of the points earned on Part A was designated as the applicant's raw score on this part. The Part A raw scores on a form were converted to a distribution having a mean of 50 and a standard deviation of 10.

Part B

The answers to each of the five questions in Part B, Form 1, were evaluated on a 20 point scale. The scores on each question were converted to a distribution having a mean of 50 and a standard deviation of 10 (so that each question would be weighted equally in determining an applicant's total score on this part). The five scale scores were summed for each applicant to obtain a total score for Part B. These total scores were converted to a distribution having a mean of 50 and a standard deviation of 10.

The answers on Part B on Forms 2 and 3 were evaluated on a 100 point scale. The Part B scores on each of these forms were then converted to a distribution having a mean of 50 and a standard deviation of 10.

Total Score

An applicant's total score on the Research Test was the average of the Part A and B scale scores.

-10-
Table 4.2

SUMMARY OF PART A ITEM ANALYSIS RESULTS

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Form 1</th>
<th>Form 2</th>
<th>Form 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Items Scored</td>
<td>42</td>
<td>40</td>
<td>42</td>
</tr>
<tr>
<td>Average Item Score</td>
<td>3.48</td>
<td>3.35</td>
<td>3.48</td>
</tr>
<tr>
<td>Average Item Variance</td>
<td>.56</td>
<td>.60</td>
<td>.56</td>
</tr>
<tr>
<td>Reliability (Alpha)</td>
<td>.87</td>
<td>.87</td>
<td>.89</td>
</tr>
</tbody>
</table>

The cases and propositions on Part A on Form 1 were exactly the same as those on Form 3. Form 2 had an entirely different set of cases and propositions.

READER CONSISTENCY IN PART B

Form 1

Table 4.3 presents the average raw score assigned by a reader to all the applicants graded by that reader on Form 1 as well as the average Part A and Phase I GBX scores of the applicants evaluated by each reader.

Table 4.3

AVERAGE SCORES OF APPLICANTS ASSIGNED TO EACH PART B READER ON FORM 1

<table>
<thead>
<tr>
<th>Question</th>
<th>Reader Number</th>
<th>Number of Applicants</th>
<th>Average Score</th>
<th>Standard Deviation</th>
<th>Average Score Part A</th>
<th>Phase I</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>228</td>
<td>8.66</td>
<td>3.69</td>
<td>145.2</td>
<td>615.9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>223</td>
<td>10.48</td>
<td>4.03</td>
<td>147.5</td>
<td>642.5</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>218</td>
<td>7.20</td>
<td>3.48</td>
<td>148.1</td>
<td>638.6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>233</td>
<td>7.38</td>
<td>3.53</td>
<td>144.8</td>
<td>619.9</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>375</td>
<td>9.79</td>
<td>4.14</td>
<td>146.1</td>
<td>630.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>76</td>
<td>7.80</td>
<td>2.92</td>
<td>147.4</td>
<td>624.1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>224</td>
<td>9.43</td>
<td>4.20</td>
<td>147.5</td>
<td>642.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>227</td>
<td>5.21</td>
<td>2.76</td>
<td>145.3</td>
<td>615.9</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>228</td>
<td>8.28</td>
<td>4.77</td>
<td>145.2</td>
<td>615.6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>223</td>
<td>4.84</td>
<td>3.24</td>
<td>147.5</td>
<td>642.6</td>
</tr>
</tbody>
</table>

It is evident from the last column of Table 4.3 that applicants were not assigned randomly to readers on each question. If random assignment had occurred, there would not not have been a large difference in the average Phase I scores between applicant groups on Questions 1, 2, 4, and 5. The disproportionate allocation of applicants to readers on Question 3 was due to administrative factors that were unrelated to the study.
COMPARABILITY OF SAMPLES

Because of the similarity between Forms 1 and 3, about 25 percent of the applicants assigned to the Research Test section of the Special Session were assigned to each of these forms. The remaining applicants were assigned to Form 2. The assignment of applicants to forms was done randomly and all test forms were given at each location at which the General Bar Examination was administered.

Table 4.1 presents summary statistical data on the sample of applicants that took each form. An inspection of these data indicates that the groups had extremely similar characteristics.

Table 4.1
SUMMARY DATA ON THE GROUP TAKING EACH FORM

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Form 1</th>
<th>Form 2</th>
<th>Form 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Anglo</td>
<td>82</td>
<td>82</td>
<td>83</td>
</tr>
<tr>
<td>Percent Male</td>
<td>64</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td>Percent from ABA schools</td>
<td>57</td>
<td>64</td>
<td>62</td>
</tr>
<tr>
<td>Percent Repeaters</td>
<td>29</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Average Age</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Average MBE score</td>
<td>423</td>
<td>424</td>
<td>424</td>
</tr>
<tr>
<td>Average Essay score</td>
<td>620</td>
<td>621</td>
<td>618</td>
</tr>
<tr>
<td>Number of Applicants</td>
<td>451</td>
<td>873</td>
<td>461</td>
</tr>
</tbody>
</table>

ITEM ANALYSIS OF PART A

A summary of the results of an analysis of the responses to each item (case/proposition combination) in Part A is presented in Table 4.2. The persons who developed the scoring key for Part A of Forms 1 and 3 were not able to achieve consensus on the ratings that should have been assigned to 3 of the 45 items in this part. Thus, only 42 items on each form were scored.

The data in Table 4.2 indicate that scores on Part A were very reliable. An inspection of the data also revealed that the score on an item in Part A correlated positively with the total score on the other items in this part (about two thirds of these correlations were over .30). This was true for all three forms of the test. These findings support the appropriateness of the scoring key.
An analysis also was conducted of the degree to which the two readers assigned to Form 1 agreed with each other in the scores that should be assigned to 20 answers to each question that were graded independently by both readers. These answers were embedded in the regular set of answers that were evaluated by each reader. In other words, each reader evaluated a set of 20 answers that were also evaluated by the other reader and these answers were embedded in each reader's regular set. The results of this analysis are presented in Table 4.4.

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Reader Number</th>
<th>Average Score</th>
<th>Standard Deviation</th>
<th>Correlation Between Readers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>10.1</td>
<td>4.2</td>
<td>.55</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9.3</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>9.1</td>
<td>4.1</td>
<td>.60</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6.7</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>13.9</td>
<td>3.7</td>
<td>.62</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>8.1</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>10.0</td>
<td>3.5</td>
<td>.67</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6.1</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>11.0</td>
<td>4.1</td>
<td>.84</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4.6</td>
<td>2.6</td>
<td></td>
</tr>
</tbody>
</table>

The correlations in the last column of Table 4.4 indicate that the two readers agreed with each other fairly well in the relative standings of the applicants (average r = .66). This level of inter-reader agreement is slightly better than that obtained on an essay question on the General Bar Examination (see Klein, 1980b).

The data in Tables 4.3 and 4.4 indicate that Reader #1 tended to assign higher grades than Reader #2. One explanation of these findings is that the low level of agreement between readers in the score that should be assigned to an answer (as distinct from the high level of agreement regarding its relative quality) probably stemmed from the readers' being more familiar with a 100 point scale than with the 20 point scale used with each Part B question on Form 1. They also could have differed in leniency.

The seriousness of the apparent differences in leniency across readers was mitigated by the considerable counterbalancing of readers and applicants across questions; e.g., if an applicant's answer to Question 1 was evaluated by Reader #1, then that applicant's answer to Question 2 was evaluated by Reader #2. Similar counterbalancing occurred on Questions 4 and 5.
Whether the counterbalancing described above was sufficient to offset the problems created by the differences in the grading standards used by the readers was investigated by scaling the scores assigned by a reader on a question to the mean and standard deviation of the Phase I scores of the applicants whose answers were graded by that reader. The average of the five scale scores constituted an applicant's "reader adjusted" Part B score. The pattern of results obtained with this score was essentially the same as that obtained with the sum of the raw scores. Thus, the results obtained with this sum (rather than with the adjusted score) are presented in the remainder of this report.

Forms 2 and 3

Forms 2 and 3 each had only one essay type question in Part B. Because of the relatively large number of applicants taking Form 2, there were three readers assigned to grade the answers on this form. Table 4.5 presents the average raw score assigned by a reader to all the Part B answers graded by that reader on a form as well as the average Part A and Phase I GBX scores of the applicants evaluated by each reader.

The differences in the average Phase I scores across reader groups, especially on Form 2, suggests that applicants were not assigned randomly to readers on this Part. However, unlike Form 1, there did not appear to be significant differences in leniency between readers on Forms 2 and 3.

Table 4.5

<table>
<thead>
<tr>
<th>Form</th>
<th>Reader Number</th>
<th>Number of Applicants</th>
<th>Average Score</th>
<th>Standard Deviation</th>
<th>Average Score Part A</th>
<th>Phase I</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>380</td>
<td>68.8</td>
<td>9.9</td>
<td>132.6</td>
<td>615.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>261</td>
<td>71.7</td>
<td>8.7</td>
<td>136.1</td>
<td>650.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>232</td>
<td>70.1</td>
<td>11.9</td>
<td>133.4</td>
<td>632.4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>225</td>
<td>67.2</td>
<td>11.0</td>
<td>146.5</td>
<td>622.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>236</td>
<td>67.2</td>
<td>13.0</td>
<td>146.2</td>
<td>636.9</td>
</tr>
</tbody>
</table>

An analysis of inter-reader agreement was conducted by embedding a common set of 20 answers in each reader's total set of answers. The average correlation between readers on Forms 2 and 3 on these answers was .75 and .65, respectively. There also was good agreement between readers in average score and standard deviation.

While there is no direct way of determining whether the absolute (as distinct from relative) quality of an applicant's performance on the Research Test was comparable to that of the applicant's performance on the GBX, it was observed that the average Part B scores on Forms 2 and 3 (67.2 and 69.7, respectively) were quite close to the average score (68.6) that was assigned to a typical question on the Essay section of the GBX. The scoring system used with Part B of form 1 precluded making a comparison between scores on this form and GBX scores.
RELIABILITY OF TOTAL SCORE

There was no direct way of assessing the reliability of the total score on the Research Test because each applicant only took one form of this test. However, the reliability of the Part A scores could be estimated since each applicant evaluated several case/proposition combinations. This reliability tended to be about .87 (see Table 4.2).

The reliability of the total Part B score on Form 1 could be estimated from the five subscores on this form. This reliability was .62 (coefficient alpha). Table 4.6 contains a summary of the item analysis results used in obtaining this estimate. An inspection of these data indicates that the score on a question correlated fairly well with the sum of the scores on the other four questions (see "Item-Total Correlations" column). The one exception to this trend occurred with question #5. This question asked the applicant to discuss whether the client should ask for a jury trial. In fact, the reliability of the total score would have increased to .68 if this question was dropped. This finding and the high reliability of the scores on this question (see Table 4.4) suggest that it may be measuring a somewhat different ability than that measured by the other questions. For example, this question called for a discussion of trial strategy whereas the other questions focused more on analytic skills.

Table 4.6
ITEM ANALYSIS RESULTS ON PART B - FORM 1

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Average Score</th>
<th>Standard Deviation</th>
<th>Item-Total Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.56</td>
<td>3.96</td>
<td>.40</td>
</tr>
<tr>
<td>2</td>
<td>7.29</td>
<td>3.50</td>
<td>.33</td>
</tr>
<tr>
<td>3</td>
<td>9.46</td>
<td>4.03</td>
<td>.50</td>
</tr>
<tr>
<td>4</td>
<td>7.31</td>
<td>4.12</td>
<td>.52</td>
</tr>
<tr>
<td>5</td>
<td>6.58</td>
<td>4.43</td>
<td>.15</td>
</tr>
</tbody>
</table>

These results along with those in Table 4.4 indicate that Part B can be scored with a reasonably high degree of reliability, especially when one considers that: (1) the reliability of three essay questions on the GBX, one test session, is only .59 (based on stepping up the average inter-item correlation of .32 within a session) and (2) the reliability of the Part B total score would have been substantially higher if the readers on this part had used the same scoring standards. In other words, if the average scores assigned by the two readers on each question were more similar, it would have increased the average correlation between questions which in turn would have increased the reliability of the total score on this part.

The reliability estimates for Part A and B (.87 and .62, respectively) were combined to estimate the reliability of the total score. The formula for estimating the reliability of a linear composite that is described by Guillicksen (1950) was used for this purpose. This formula indicated that the total score on the Research Test had a reliability of about .78. This is almost as high as the reliability of the entire Essay section of the GBX.
Chapter 5

RESULTS

RELATIONSHIP BETWEEN RESEARCH AND GBX SCORES

Table 5.1 presents the correlation among the Part A (ratings of case/proposition combinations), Part B (memoranda), and Total Research Test scores as well as their correlations with the MBE and Essay portions of the General Bar Examination (GBX). The data with Form 3 were combined with that of Form 1 for this table because Part A was the same on these two forms and the correlations with their respective Part B scores were almost identical on every measure examined. The correlations for Forms 1+3 appear above the main diagonal and for Form 2 below the main diagonal.

Table 5.1

CORRELATIONS AMONG RESEARCH TEST AND GBX SCORES

<table>
<thead>
<tr>
<th>RESEARCH SCORES</th>
<th>GBX SCORES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Part A</td>
</tr>
<tr>
<td>Part A</td>
<td>--</td>
</tr>
<tr>
<td>Part B</td>
<td>.33</td>
</tr>
<tr>
<td>Total</td>
<td>.81</td>
</tr>
<tr>
<td>MBE</td>
<td>.38</td>
</tr>
<tr>
<td>Essay</td>
<td>.32</td>
</tr>
<tr>
<td>GBX Total</td>
<td>.38</td>
</tr>
</tbody>
</table>

The data in Table 5.1 indicate that there was a low positive correlation between the Part A and B scores on each Form (.35 and .33 for Forms 1+3 and 2, respectively). These correlations are about as strong as the correlation between two essay questions on the GBX (average \( r = .32 \), see Klein, 1981a). The Part A scores on Forms 1 and 3 were more highly correlated with MBE scores than they were with Essay scores while the reverse was true for Part B scores. There was a .60 average correlation between the total scale score on the Research Test and the total GBX score. This relationship is almost as strong as the one between GBX Essay and MBE scores (\( r = .69 \)). The very high correlations between a part score and a total score on the same test were, of course, a result of including the part scores in the computation of the total scores.

RELATIONSHIP OF RESEARCH TEST SCORES TO OTHER MEASURES

Table 5.2 presents the correlation of the Part A, Part B, and Total Research Test scores with other measures. For the purposes of comparison, the last two columns of this table present the correlation between each measure and scores on the MBE and Essay portions of the GBX for all the applicants who took one of the three forms of the Research Test.
Table 5.2

RELATIONSHIP OF RESEARCH TEST AND GBX SCORES TO OTHER MEASURES

<table>
<thead>
<tr>
<th>Measure</th>
<th>RESEARCH TEST SCORES</th>
<th></th>
<th></th>
<th></th>
<th>GBX SCORES</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1+3 2</td>
<td>1+3 2</td>
<td>1+3 2</td>
<td>1+3 2</td>
<td>MBE Essay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-.35 -.31</td>
<td>-.28 -.38</td>
<td>-.38 -.38</td>
<td>-.29 -.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Being Male</td>
<td>-.03 .00</td>
<td>-.18 -.09</td>
<td>-.10 -.05</td>
<td>.00 -.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Being Anglo</td>
<td>.20 .18</td>
<td>.17 .18</td>
<td>.27 .21</td>
<td>.26 .24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Being Hispanic</td>
<td>-.07 -.05</td>
<td>-.06 -.12</td>
<td>-.09 -.12</td>
<td>-.13 -.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Being Black</td>
<td>-.12 -.18</td>
<td>-.16 -.14</td>
<td>-.19 -.14</td>
<td>-.19 -.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graduate of a CA ABA Law School</td>
<td>.21 .21</td>
<td>.19 .27</td>
<td>.25 .29</td>
<td>.25 .27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Being a Repeater</td>
<td>-.23 -.31</td>
<td>-.33 -.41</td>
<td>-.40 -.44</td>
<td>-.45 -.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worked as a Law Clerk</td>
<td>.18 .23</td>
<td>.22 .31</td>
<td>.22 .33</td>
<td>.23 .28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Received pay for legal research</td>
<td>.20 .25</td>
<td>.26 .33</td>
<td>.28 .34</td>
<td>.26 .28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hours of paid legal research</td>
<td>.16 .23</td>
<td>.23 .24</td>
<td>.22 .27</td>
<td>.20 .21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The correlation coefficients in Table 5.2 indicate that Research Test scores correlate with other measures in the same direction and to about the same degree as GBX scores. For example, there was a -.42 correlation between being a repeater and total Research scores as well as between being a repeater and GBX scores. Similarly, there was a .28 correlation between being a graduate of a California ABA approved law school and Research Test scores and a .26 correlation between being an ABA graduate and GBX scores.

The Research Test did not significantly narrow the gap in average performance level between Anglo and minority groups (Klein, 1979). For example, the typical Anglo applicant had a GBX score at the 55th percentile and a Research Test score at the 54th percentile. The typical minority applicant had GBX and Research Test scores at the 28th and 30th percentiles, respectively. The 3 point difference in the size of these gaps was most likely due to the Research Test having a slightly lower reliability than the GBX.

Only three questionnaire items were consistently and statistically significantly correlated with Research Test scores; namely: worked as a law clerk, paid for doing legal research, and the number of hours of legal research for which the applicant was paid. However, these variables also were correlated with GBX scores. And, all the correlations among them could stem from some other variable. For example, the applicants who are paid to do legal research may be the ones who tend to have high law school grades. Thus, engaging in these activities may not have helped an applicant achieve higher GBX or Research Test scores than would otherwise be the case.
Rerunning all the correlations with the data from just first time takers (so as to control in part for the amount of time applicants had for engaging in various types of activities and training) did not appreciably affect the relationships presented in Tables 5.1 and 5.2.

ATTITUDES TOWARDS THE TEST

The evaluation forms the applicants completed on the third day of the GBX were analyzed to assess applicant attitudes towards the Research Test. Table 5.3 contains the average score on each characteristic evaluated and the percentage of applicants who rated each test as Good or Very Good on each dimension. It is evident from this table that the applicants judged the Research Test to be a more realistic and a better measure of their ability to perform as an attorney than either the Essay and MBE portions of the GBX. However, they rated the Essay and MBE as better indicators of their legal knowledge.

Table 5.3

<table>
<thead>
<tr>
<th>Characteristic Evaluated</th>
<th>Test Evaluated</th>
<th>Average Score in Group that Took Form 1+3</th>
<th>Average Score in Group that Took Form 2</th>
<th>Percent Rating it Good or Very Good in Group that Took Form 1+3</th>
<th>Percent Rating it Good or Very Good in Group that Took Form 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal Knowledge</td>
<td>Research</td>
<td>2.8</td>
<td>2.4</td>
<td>25</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Essay</td>
<td>3.4</td>
<td>3.4</td>
<td>48</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>MBE</td>
<td>3.1</td>
<td>3.1</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td>Ability to Perform as an Attorney</td>
<td>Research</td>
<td>3.0</td>
<td>2.5</td>
<td>35</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Essay</td>
<td>2.8</td>
<td>2.8</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>MBE</td>
<td>2.3</td>
<td>2.3</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Realism of Case Situations</td>
<td>Research</td>
<td>3.9</td>
<td>3.7</td>
<td>76</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Essay</td>
<td>3.0</td>
<td>3.1</td>
<td>33</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>MBE</td>
<td>2.7</td>
<td>2.8</td>
<td>23</td>
<td>26</td>
</tr>
</tbody>
</table>

The rating scale for the items was 1=very poor, 2=poor, 3=fair, 4= good, and 5=very good.

The most frequent complaint about the Research Test was that there was not enough time to complete it. For example, 84 percent of the applicants made this complaint about the test while only 27 percent said the time limits on the Essay and MBE portions of the examination were too short.

There was no systematic relationship between an applicant's evaluation of the quality of the Research Test (including the appropriateness of its time limits) and that applicant's score on any of the parts of this test. Similarly, there was no systematic relationship between an applicant's evaluation of the quality of the MBE and Essay sections of the GBX and that applicant's scores on these sections. All the correlations were less than .08.
Chapter 6

SUMMARY AND IMPLICATIONS OF FINDINGS

The results presented in Chapters 4 and 5 indicate:

- An applicant's performance on Part A of the Research Test was quite reliable. This suggests that an appropriate scoring key can be developed for this type of measure and that an applicant's ability to evaluate the degree to which one case supports a given legal proposition is related to how well that applicant can evaluate other case/proposition combinations. It also suggests that an applicant's workload on the Research Test can be eased somewhat by reducing the number of case/proposition combinations the applicant has to evaluate.

- An applicant's performance on Part B of the Research Test can be scored at least as reliably as an answer written on the Essay section of the GBX. However, more effort will have to be devoted to reader calibration on Part B if scores are assigned using something other than the traditional 100 point scale.

- Like the MBE and Essay sections of the GBX, the two parts of the Research Test measure similar but not identical skills. Thus, it may not be appropriate to reduce scoring costs and testing time on the Research Test by replacing Part B with a machine scoreable version of Part A. Similarly, it would not be appropriate to drop Part A, especially since it contributes substantially to the reliability of the Research Test's total score.

- Total scores on the Research Test correlated with total GBX scores only slightly less well than MBE and Essay scores correlated with each other. While this result may be due in part to the Research Test scores not being quite as reliable as the MBE and Essay scores, it also may stem from the Research Test measuring skills and knowledge that are similar but not identical to those that are now assessed by the GBX.

- The Research Test did not widen or narrow the gap in performance levels between racial groups that is present on the GBX.

- Repeater status, type of law school from which the applicant graduated, and most other background characteristics were just as highly correlated with Research Test scores as they were with scores on the GBX. For example, applicants who had worked as law clerks and received pay for doing legal research prior to taking the Research Test tended to have higher scores on this test (and on the GBX) than applicants who did not have this training. These findings do not necessarily mean that such experience caused the higher test scores in that the applicants who are selected for these jobs also may be the ones with higher than average law school grades.

- There was no relationship between an applicant's assessment of the quality of the Research Test, MBE, or Essay sections of the examination and that applicant's scores on these sections.
Questionnaire data did indicate that applicants thought the Essay and MBE portions of the GBX were better measures of their legal knowledge than was the Research Test. However, they thought the Research Test was a better measure of their ability to perform as an attorney. They also felt that the case situations presented in the Research Test were far more realistic than those used in either the Essay or MBE sections of the GBX.

The applicants' chief complaint about the Research Test was that they did not have enough time to complete it.

The pattern of relationships with scores on one form of the Research Test was essentially the same as the pattern of relationships with scores on the other two forms of the test. This finding suggests that the general test plan used to construct these forms could be used to develop additional forms of the test that would have essentially the same characteristics as the forms used in this study.

While there is no direct way of determining whether the absolute (as distinct from relative) quality of an applicant's performance on the Research Test was comparable to that applicant's performance on the GBX, it was observed that the average Part B scores on Forms 2 and 3 were almost the same as the average score that was assigned to a typical question on the Essay section of the GBX.

The foregoing results suggest that the Research Test measures important skills that are similar but not identical to those measured by the MBE and Essay portions of the GBX. In other words, it appears that the Research Test provides unique information about an applicant's legal abilities. And, there is consensus in the legal profession that these abilities are material to the practice of law. Thus, adding the Research Test to the GBX could increase the validity of the bar examination process.

One major obstacle to adding a Research Test to the bar examination is the cost of developing a new form of the test for each administration of the GBX. Unlike the typical essay question which is only a few paragraphs long, the Research Test requires the production and duplication of over 50 pages of case materials and other documents to be evaluated by each applicant. While it appears that an applicant's workload on the Research Test can and should be reduced, each new form of the test will still probably take longer (and cost more) to develop than the entire Essay section of the GBX. Thus, California may want to explore the feasibility of spreading out the test development costs by joining with other jurisdictions in the test construction process in much the same way that the MBE's development costs are shared by the jurisdictions that use the MBE.

Finally, it was not possible within the constraints of the July 1980 bar examination to have a given applicant take more than one form of the Research Test. Thus, there was no way of assessing directly the degree to which the scores on the test were affected by the unique nature of the case materials in each form. It is recommended, therefore, that this matter be investigated further so as to determine the reliability of the scores obtained. Such a study might be conducted in cooperation with two or three law schools.

-20-
REFERENCES


APPENDIX A

EXCERPTS FROM NOTICE TO APPLICANTS REGARDING THE SPECIAL SESSION

In addition to the usual examination sessions, each applicant for the Fall 1980 Examination will be expected to participate in a one-half day Special Session to be administered either on Monday, July 28, 1980, or Tuesday, July 29, 1980.

The Special Session is designed to give applicants an opportunity to improve their scores by demonstrating their capabilities in other than the standard methods or procedures and to test alternative means of examining for the future.

To derive any benefit from the Special Session, applicants must complete and return designated questionnaires by specified deadlines and complete all other portions of the examination which they are taking.

There will be four types of problems given during the Special Session. Individual applicants will be assigned to the various types of problems by the Committee based on a statistical analysis of the composition of the entire examinee population. General descriptions of the four types of problems are:

Research Problem

Applicants will be given factual and legal information regarding a possible case for a client. They will then be asked to evaluate the extent to which certain legal materials given them support or are contrary to various legal propositions involved in the client's case. Applicants will also be asked to prepare a memorandum regarding various aspects of the client's case. Thus, this problem will be almost like an "open-book" test.

The scoring for the Special Session for the various groups of applicants will be scaled so that the Special Session would be equivalent to one session of a six-session examination. An applicant's score on the Special Session will then be utilized if and only if (a) the applicant did not pass the regular part of the examination based upon the usual scoring procedures and (b) the applicant's score on the Special Session is above passing. Thus, an applicant's performance on the Special Session can increase but cannot decrease the likelihood that the applicant will pass the examination. Since the score on the Special Session can count for one-sixth of an applicant's total grade, it can have a substantial beneficial impact for an applicant who scores well on it.
APPLICATION - AUGUST 1980 ASSESSMENT CENTER AND QUESTIONNAIRE FOR ALL JULY, 1980 APPLICANTS

Your answers to the questions below will assist the Committee of Bar Examiners in its efforts to improve the bar examination. Your responses will be kept strictly confidential and used solely for statistical purposes. We are most appreciative of your cooperation.

LAST NAME ___________________________ FIRST NAME ___________________________ MID INITIAL

1. Birthdate [ ] / [ ] / [ ]
   Mo. Day Yr.

2. Which of the following jobs, if any, have you held for one or more months? Check all that apply:
   [ ] Attorney
   [ ] Court Reporter
   [ ] Law Clerk
   [ ] Law Enforcement Officer
   [ ] Legal Investigator
   [ ] Legal Secretary
   [ ] Paralegal Assistant
   [ ] Other Law Related Employment (specify____________________)

3. About what percentage of the time was English spoken in your home during your childhood? Please put your answer in the box below.
   [ ] Percent of time English spoken.

4. Circle the box corresponding to the language other than English that was spoken most often.
   [1] Chinese
   [3] Spanish
   [4] Tagalog/Phillipino
   [5] None of these or no language besides English

5. Circle the box corresponding to the highest grade in school completed by your MOTHER or female guardian. Leave blank if you do not know.
   [1] 1st to 5th
   [2] 6th to 8th
   [3] 9th to 11th
   [4] High School Graduate
   [5] College Graduate
   [6] Graduate/Professional School Degree

6. Circle the box corresponding to the highest grade in school completed by your FATHER or male guardian. Leave blank if you do not know.
   [1] 1st to 5th
   [2] 6th to 8th
   [3] 9th to 11th
   [4] High School Graduate
   [5] College Graduate
   [6] Graduate/Professional School Degree

PLEASE CONTINUE ON THE REVERSE SIDE
7. Circle the box below that best describes your undergraduate major:
   [1] Economics, Business, Accounting
   [2] Physical Sciences, Engineering, Mathematics, Biology
   [3] Social Sciences (Anthropology, Psychology, Sociology)
   [5] English, Journalism, Classical Studies, Philosophy
   [7] Education
   [8] Other (specify) ..................................................

8. For each of the activities below, indicate the number of hours, if any, you have spent doing them for:
   SIMULATED cases, ACTUAL cases associated with supervised law school programs, and ACTUAL cases as
   part of paid and/or volunteer employment. Insert the number 99 in a box if you spent 100 or more hours
   doing the activity.

   ACTUAL CASES
   Supervised Law School
   Paid or Volunteer
   Cases Programs Employment

   a. Conduct legal research. [ ] [ ] [ ]
   b. Prepare briefs, petitions, or motions. [ ] [ ] [ ]
   c. Conduct direct examinations. [ ] [ ] [ ]
   d. Conduct cross examinations. [ ] [ ] [ ]
   e. Interview a client or a witness for a hearing. [ ] [ ] [ ]
   f. Interview a client on general legal matters; e.g., landlord-tenant dispute. [ ] [ ] [ ]
   g. Present an oral argument in a legal proceeding. [ ] [ ] [ ]

9. In the box next to each choice below, indicate the number of courses, if any, you have taken in:
   [ ] Evidence
   [ ] Clinical or Trial Practice involving your participation in simulated or actual hearings
   [ ] Trial Practice NOT involving your participation in simulated or actual hearings; i.e., lecture only.

10. If you have secured employment in California in a law related job commencing by September 15, please
    circle the box corresponding to how you will be employed.
    [1] Attorney General's Office
    [2] Public office, criminal prosecution (e.g., District Attorney)
    [3] Public office, criminal defense (e.g., Public Defender)
    [4] Legal aid office (e.g., neighborhood legal assistance)
    [5] Public interest law firm
    [6] Law clerk for a judge
    [7] Private law firm, criminal defense
    [8] Private law firm, general practice
    [9] Law department of a corporation
    [10] Other (specify) ..................................................

11. Do you hereby apply to participate in the two day Assessment Center? (Circle your answer below.)
    (1) No
    (2) Yes

If selected as a participant in the Assessment Center, I will be available to participate in the location circled
below on any date between August 4 and August 24 (inclusive) except for the dates circled below.

Circle one

   in (1) Los Angeles only
   in (2) San Francisco only
   in (3) Either Los Angeles or San Francisco.

   Mon Tue Wed Thu Fri Sat Sun
   4  5  6  7  8  9  10
   11 12 13 14 15 16 17
   18 19 20 21 22 23 24

(Note: The likelihood of being selected may be improved by increased availability.)

Date: ______________, 1980  Signature: __________________________
APPENDIX C
POST EXAMINATION QUESTIONNAIRE

APPLICANTS MUST COMPLETE AND FILE THIS QUESTIONNAIRE BEFORE THE FINAL SESSION OF THE GENERAL BAR EXAMINATION ON WEDNESDAY, JULY 30, 1980 IN ORDER TO DERIVE ANY BENEFIT FROM THE SPECIAL SESSION OF THE JULY 1980 EXAMINATION OR TO PARTICIPATE IN THE AUGUST 1980 ASSESSMENT CENTER.

Your answers to the questions below will assist the Committee of Bar Examiners in its efforts to improve the bar examination. Your responses will be kept strictly confidential and used solely for statistical purposes. A completed questionnaire is required of all applicants who participated in any one of the special sessions and/or will be participating in the Assessment Center. We are most appreciative of your cooperation in answering the questions below.

Application # [ ] Center Number [ ] Birthdate [ / / ]

1. Which of the following methods, if any, did you use to prepare for the examination? Check all that apply:

[ ] Commercial bar review course of at least 5 weeks duration which met 4 or more times per week; e.g., BAR or BRC.
[ ] Commercial bar review course of less than 5 weeks duration.
[ ] Intensified commercial writing course; e.g., Beverly Rubens.
[ ] Law school sponsored or administered bar review course.

2. About how many hours per week were you engaged in paid employment during the last five weeks? Please put your answer in the box below.

[ ] Hours per week Employed

3. About how many hours per week did you spend studying for the examination during the last five weeks? Please put your answer in the box below.

[ ] Hours per week Studied

4. Circle the number corresponding to the special session you took:

1 Videotape of Arbitration and Court Room cases
2 Case of Barelas v. Brown (Research Task)
3 Case of State v. Dolan (Research Task)
4 60 multiple choice questions and script with questions
5 Two essay questions and 20 multiple choice questions
9 Did not participate in a special session

5. In your opinion, how good a measure of your LEGAL KNOWLEDGE was each part of the examination? Please circle one number for each part.

<table>
<thead>
<tr>
<th></th>
<th>Very Poor</th>
<th>Poor</th>
<th>Fair</th>
<th>Good</th>
<th>Very Good</th>
<th>Did Not Take</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Special Session</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>b. Essay Examination</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>c. Multistate (MBE)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>

-25-
6. In your opinion, how good a measure of your ABILITY TO PERFORM AS AN ATTORNEY was each part of the examination? Please circle one number for each part.

<table>
<thead>
<tr>
<th></th>
<th>Very Poor</th>
<th>Poor</th>
<th>Fair</th>
<th>Good</th>
<th>Very Good</th>
<th>Did Not Take</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Special Session</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>b. Essay Examination</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>c. Multistate (MBE)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>

7. In your opinion, was the time allowed for each part the examination appropriate? Please circle one number for each part.

<table>
<thead>
<tr>
<th></th>
<th>Less Than Enough</th>
<th>About Right</th>
<th>More Than Enough</th>
<th>Did Not Take</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Special Session</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>b. Essay Examination</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>c. Multistate (MBE)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>9</td>
</tr>
</tbody>
</table>

8. Please circle one number below for each part of the examination to indicate how much you agree or disagree with the statement: "The case situations were realistic."

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Did Not Take</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Special Session</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>b. Essay Examination</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>c. Multistate (MBE)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>

9. In general, were the directions and questions in your special sessions clear or ambiguous? Please circle one choice below.

1-Clear  2-Mixed  3-Ambiguous  9-Did Not Take

10. What were your general reactions to the special session? How could it be improved? Should it be included in future examinations?