

AN ANALYSIS OF THE RELATIONSHIP BETWEEN
TRIAL PRACTICE SKILLS AND BAR EXAMINATION RESULTS

A Report Prepared for 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

January 10, 1983

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 conduct legal research and to make timely and appropriate decisions during a trial). These tests were given in conjunction with the July 1980 administration of California's general bar examination (GBX).

The results of these studies are being presented in a series of reports. The first report (Klein, 1981a) discussed the findings with the Research Test. As its name implies, the Research Test assessed 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.

The second report presented the results of an Assessment Center (AC) that was conducted with 485 applicants who also took the July 1980 GBX (Klein and Bolus, 1982). The AC was conducted over a two-week period in August of 1980. Each AC participant attended for two days. On one day, the participant served as counsel for the plaintiff in a simulated case and on the other day, as counsel for the defendant in a different simulated case. On both days, the participant took several oral and written tasks. Analyses of the data examined the relationships among the different types of AC scores and their relationships with scores on the GBX and other measures. Special attention was given to investigating whether the inclusion of AC type tasks on the GBX would affect minority passing rates.

This report describes the procedures that were used and the results obtained with two tests that were designed to assess certain trial practice skills. In each of these 90 minute tests, applicants were given background information about a case that was in trial or arbitration. They were then shown a videotape of a brief segment of the proceedings. This was followed by one to three questions about the segment, such as whether a certain objection was valid. These questions appeared on the TV monitor and the applicants had 5 to 10 minutes to write their answers before the next segment appeared on the screen. There were 11 segments per test.

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 them. The results of these studies are presented in another report (Klein, 1981b).

SUMMARY

Two forms of a Trial Practices Test (TPT), designated A and B, were developed to provide measures of certain important lawyering skills that were not tested (at least not directly) by California's General Bar Examination (GBX). One unique feature of the TPT was that applicants had to answer 1 to 3 specific questions about a simulated arbitration (Form A) or trial (Form B) immediately after viewing a short segment of the proceedings on a television monitor. Each form took about 90 minutes to complete and contained 11 segments (19 scorable questions).

Two types of questions, Content and Strategy, were usually asked after each segment. Content questions generally inquired about the validity of objections, such as "discuss how the judge should rule on the objection." Strategy questions focused on the appropriateness of trial tactics, such as "should the plaintiff have attempted to elicit the information on direct examination?"

The 323 applicants at the San Francisco test center took the forms in the sequence A then B whereas the 538 applicants at the Long Beach center took them in the sequence B then A.

The results of analyses with TPT scores indicated:

- o TPT answers were scored just as reliably as Essay answers on the GBX. The scores on one half day of TPT tests were just as reliable as the scores on one full day of the GBX Essay.
- o How well an applicant performed on one form of the TPT was a good predictor of how well the applicant performed on the other form (observed and corrected correlations were .58 and .90, respectively).
- o The sequence in which the forms were taken had no practical impact on scores.
- o Although the TPT included both Content and Strategy questions, there was no distinction between these types in terms of applicant performance. An applicant's score on one type of question was just as highly correlated with scores on other questions of the same type as with scores on questions of a different type.
- o The TPT was just as highly correlated with the Essay and MBE as the Essay and MBE were correlated with each other. Thus, the TPT was measuring skills that were similar to but not exactly the same as those assessed by the GBX.
- o TPT and GBX scores had very similar correlations with various applicant background characteristics, such as race, sex, age, repeater status, and law school type. Thus, it did not widen or narrow the gap in average scores that currently exists among groups on the GBX.

- o Applicants who had some clinical experience did better on the TPT than those who had no experience. This difference was not attributable to differences in average GBX scores among experience groups. There also was a strong underlying relationship between TPT and Assessment Center scores.
- o Although applicants thought the GBX was a better test of their legal knowledge than the TPT, they had the opposite opinion regarding the relative efficacy of the measures as tests of an applicant's ability to perform as an attorney. They also felt the TPT presented more realistic case situations than the GBX.
- o Average scores on TPT answers were related to presentation mode (video versus scripted), but the nature of this relationship suggested that it may have been produced by factors that were related to the experimental procedures rather than to the mode per se. Nevertheless, it appears that the two modes are likely to yield similar results.

The foregoing findings suggest that including a TPT in the GBX would increase the GBX's reliability, make a small but noticeable change in who passed, and increase the validity of the examination by measuring skills that are relevant to practice but which are not assessed now.

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 Leo O'Brien of Hastings Law School was primarily responsible for developing the test materials used in this research. Allison Hoffman and Terry Lamb supervised the production of the videotapes and their presentation during the examination.

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 answers. 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.

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..... i

SUMMARY..... ii

ACKNOWLEDGMENTS..... iv

Chapter

1. INTRODUCTION..... 1

 Background..... 1

 Purpose..... 1

 General Procedures..... 2

 Special Session..... 3

 Background Characteristics..... 3

 Preexamination Questionnaire..... 4

 Post Special Session Questionnaire..... 4

2. TEST DESCRIPTION AND PRELIMINARY ANALYSES..... 5

 Overview..... 5

 Trial Practices Test..... 5

 Test Administration..... 6

 Sample Characteristics..... 7

 Scoring..... 7

 Interreader Reliability..... 8

 Sequence Effects..... 9

 Item Analyses and Test Reliability..... 9

 Factor Analyses..... 10

 Subscore Analyses..... 10

3. RESULTS..... 12

 Overview..... 12

 Relationship with GBX Scores..... 12

 Relationship with Background Characteristics..... 12

 Relationship with Clinical Experience and Skills..... 13

 Applicant Evaluations of TPT..... 13

 Presentation Mode Study..... 14

 Summary of Results and Conclusions..... 16

Appendixes

A. Excerpts From Notice to Applicants..... 17

B. Preexamination Questionnaire..... 18

C. Post Special Session Questionnaire..... 20

REFERENCES..... 22

Chapter 1

INTRODUCTION

BACKGROUND

The general goal of a bar examination is to assess the degree to which examinees have some of the important skills and knowledge that are necessary for legal practice. Most bar examinations are therefore designed to emphasize understanding of basic legal principles and legal analysis. One skill that is not measured, but which is important for many lawyers' practices, is the ability to represent a client's interests effectively during a trial or hearing.

We do not know whether the scores on present bar examinations provide a good indicator of trial practice skills. We also do not know whether applicants to the bar who have had trial experience, such as through a clinical education program, would be more proficient in performing actual trial tasks than those who have not had such training.

One important reason we do not know more about the relationship between bar examination scores and trial practice skills is that there are a paucity adequate measures of these skills. Although trial practice exercises have been used in educational programs for both law students and attorneys, these tasks are not truly appropriate for a bar examination. They cannot be given simultaneously to large groups of applicants under standardized and secure testing conditions. There is also a question about whether performance on many of these exercises can be evaluated reliably.

PURPOSE

The foregoing considerations led the National Conference of Bar Examiners (NCBE) and the Committee of Bar Examiners (CBE) of the State Bar of California to develop a group administered test that was designed to assess certain trial practice skills. The studies presented in this report investigated whether responses on this test:

- o could be scored reliably;
- o were more a function of applicant skills than the unique characteristics of a particular version of the test;
- o were affected by how much practice applicants had in taking such tests (e.g., possible warm up or sequence effects);
- o were related to the applicants' bar examination scores, prior legal training and experience, performance on other clinical skills tasks, and various background characteristics; and
- o were influenced by whether the test materials were presented in written form or via a videotape of a hearing.

We also investigated applicant attitudes towards the test, such as whether they felt it was a good measure of their legal skills and abilities.

GENERAL PROCEDURES

All the applicants who participated in this research took the July 1980 version of California's General Bar Examination (GBX). This version of the GBX had two sections, the multistate bar examination (MBE) and the Essay.

The MBE consisted of 200 multiple choice questions (or items) that were drawn from six content areas, such as torts and contracts. Applicants were given 3 hours to answer the first 100 items and another 3 hours to answer the remaining 100 items. Raw scores on the MBE (i.e., the number of items answered correctly) were adjusted by the Educational Testing Service for possible differences in average item difficulty across administrations. California multiplied the adjusted scores by 3.0 so that the maximum possible MBE score was 600.

The Essay portion of the July 1980 GBX was divided into three sessions. In each session, applicants were given 3 hours to answer 3 questions. An applicant's answer to a question was graded on a scale of 0 to 100.

All the applicants did not have all of their essay answers graded. Instead, applicants were divided into two groups on the basis of their score on the combination of three essay questions and the MBE (hereafter referred to as the Phase I score). If the Phase I score was greater than 665, the applicant passed. If the score was not greater than 665, all of the applicant's answers were read at least once and as many as three times. Applicants needed a total score (MBE + 9 question Essay) of 1050 in order to pass if all of their answers were graded (see Klein, 1980a for a more complete description the pass/fail rules).

Table 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 ($r = .32$). 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 .68 correlation between the MBE and Essay sections.

Table 1

SUMMARY STATISTICAL DATA ON THE JULY 1980 GBX

<u>Statistic</u>	<u>MBE</u>	<u>Essay</u>	<u>Total</u>
Average Score	424.5	617.5	1042.0
Standard Deviation	46.0	54.4	92.2
Reliability	.88	.81	.91
<hr/>			
Total Score = MBE + (9)(average essay question score)			

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, 1981c).

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 as well as each section's administration and scoring costs. The subsequent chapters of this report discuss the data obtained in two sections of the SS.

BACKGROUND CHARACTERISTICS

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.

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.

PREEXAMINATION QUESTIONNAIRE

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.

POST SPECIAL SESSION QUESTIONNAIRE

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 2

TEST DESCRIPTION AND PRELIMINARY ANALYSES

OVERVIEW

Two versions (forms) of a Trial Practices Test were developed. Each form was designed to measure the degree to which applicants had some (but certainly not all) the skills that are necessary for representing a client during a hearing. One unique feature of this test was that the applicants had to answer specific questions about a simulated trial or arbitration, such as whether a particular objection was valid, immediately after viewing a short segment of the proceedings on a TV monitor.

Both versions of the test were given to two groups of applicants. In addition, a portion of the script for one of the forms and some its questions were given to two of the four groups that participated in the MBE Time Limits Study (Klein, 1981b). The other two groups in the MBE Time Limits Study received a portion of the script for the other form and some its questions. The answers written by applicants who received the scripted materials were merged with those who viewed the videotapes prior to the grading process. Graders were not informed whether a given answer was written by an applicant who had viewed a segment or read the script for it.

The remainder of this chapter discusses the characteristics of each form of the test and the results of preliminary analyses that were run on the data collected with them.

TRIAL PRACTICES TEST

Two forms of the Trial Practices Test (TPT) were developed. Form A involved an arbitration of a life insurance claim in which the beneficiary claimed the death of a spouse was accidental and the insurer asserted it was suicide. Form B involved an automobile accident in which the issue of liability was being tried before a jury. Both forms had the following features:

- o Applicants were instructed that the Federal Rules of Evidence applied. Objections of "immateriality" and "irrelevance" were to be treated as exact equivalents. They were not to discuss common law distinctions between them; but to assume that either objection sufficed to raise either point.
- o Prior to playing the tape, applicants were given background documents related to the case, such as a police report, a witness' statement, and a portion of the application for an insurance policy. Applicants were given about 10 minutes to review these materials.
- o Applicants also were provided with an answer book that contained one page for each question and an excerpt from the Federal Rules of Evidence.

- o In the first segment of each tape, the announcer introduced the judge and the parties for each side. This segment also contained a portion of the direct examination of one witness. Following this segment, two questions appeared on the screen. The announcer then advised the applicants that the segment would be played again and that the same set of questions would appear. However, he further advised them that this was the only segment that would be repeated.
- o The pattern of a brief (3 to 7 minute) segment followed by 1 to 3 questions was repeated for 10 more segments. Some of these segments involved direct examination of a witness whereas others involved cross-examination. The announcer read the questions aloud once when they first appeared on the screen.
- o Two types of questions were usually asked after each segment.

Content questions generally inquired about the validity of objections, such as "discuss how the judge should rule on the objection."

Strategy questions focused on the appropriateness of trial tactics, such as "should the plaintiff have attempted to elicit the information on direct examination?"

- o Applicants were given 5 to 10 minutes to write their answers. The amount of time allowed on a segments' questions was based on the number of questions asked and test developers' judgments regarding their difficulty. The questions appeared on the screen throughout this writing period along with the amount of time that remained before the next segment began. Thus, time limits were maintained by the videotape rather than by a test administrator.
- o Applicants were not given feedback about the adequacy of their answers to the questions on one segment by the events that took place in subsequent segments.

TEST ADMINISTRATION

It took about 90 minutes to administer each video form. Applicants were given a 15 minute break between forms. The applicants at the San Francisco test center took the forms in the sequence A then B whereas Long Beach applicants took them in the sequence B then A.

There were hardly any complaints about the ability to see the TV monitors or read the questions that followed each segment. However, several minor administrative and technical problems were encountered at both test centers. For instance, there was a partial loss of sound on one segment of Form A at Long Beach, there was a significant error in the wording of one question (that ultimately led to it not being scored), the noise from the shuffling of papers made it difficult to hear the voices on the tape on the first segment (this was quickly corrected by turning up the sound), and many applicants wrote the answers to all of the first segment's questions on the answer sheet reserved for just question #1.

SAMPLE CHARACTERISTICS

Table 2 presents summary statistical information on the applicants who took the test. These data indicate that the San Francisco applicants generally had higher scores on the GBX than the Long Beach applicants. Minority group applicants also made up a larger proportion of the Long Beach group. A comparison of the means in Table 2 with those in Table 1 indicate that the total sample taking the test had only slightly lower GBX scores than the entire population of July 1980 applicants.

Table 2

AVERAGE GBX SCORES AND PERCENTAGE OF
MINORITY APPLICANTS AT EACH TEST CENTER

Variable	San Francisco	Long Beach	Total sample
MBE	434	412	420
Essay	633	601	613
Total	1067	1013	1033
Number of applicants	323	538	861
% minority	15	25	21

SCORING

A total of 19 questions were scored on each form (i.e., not counting the question that was dropped from Form A).

Two to three readers were assigned to each question. A reader usually graded answers to all the questions on a segment because an applicant's answer to one of these questions often alluded to information contained in his/her answer to another question on that segment. However, most readers graded the answers to questions on only two or three segments.

The readers assigned to a question reviewed the segment and background materials on which the question was based and discussed the appropriate answer(s) to the question. Each reader then independently evaluated a small sample of answers that also had been graded by the other reader(s) assigned to the question. Scoring discrepancies were discussed and additional answers read until the readers appeared to be calibrated.

A five point relative grading scale was used for each question (ranging from 1 = very poor response to 5 = very good response).

INTERREADER RELIABILITY

Interreader reliability refers to the degree to which different graders independently assign the same score to a given answer. The data for the interreader reliability analyses were obtained by embedding a common set of 30 answers into the entire batch of answers assigned to each reader. Although the readers were advised that their grading would be checked in this way, they did not know which answers would be evaluated by other readers (i.e., as distinct from the calibration session where the readers knew that all the answers they were grading would be scored by others).

Analyses of the scores on the embedded answers that were evaluated by more than one reader indicated:

- o With few exceptions, readers spread their scores out around their average score in about the same way (on both forms, the mean standard deviation on a question was about 1.0 and the average range of standard deviations on a question was 0.3).
- o One reader was usually not much more lenient than another reader. The average difference in mean score between the most and least lenient reader on a question was 0.3 on Form A and 0.6 on Form B. A few questions on Form B exhibited differences in mean scores between readers that were greater than one standard deviation.
- o The average correlation between two readers was .58 on a Form A question and .44 on a Form B question. These values are typical of those obtained between two readers of a regular GBX essay question (Klein, 1977 and 1980). For instance, on the July 1979 GBX, the average correlation between two readers was .48.

An in depth analysis of some of the unusually large differences among readers on Form B revealed there was one reader who had a negative correlation with the other readers assigned to question #7 and near zero correlations with the readers on another question. These data along with anecdotal reports suggested this reader may have reversed the grading scale for all or many of the answers on question #7 and reversed it for some (but not all) of the answers on another question (in other words, he may have assigned 1's to the best answers and 5's to the worst answers on at least one of the questions).

The problem of potential score reversals was investigated further by examining the correlation between the scores a reader assigned to all the answers he/she graded on a question and these applicants' GBX scores. This analysis did not provide conclusive evidence of score reversals (because the variation among readers in the correlations between TPT question and GBX scores could easily have occurred by chance).

Various techniques were used in an attempt to adjust the readers' scores for some of the problems noted above. The results with these procedures suggested the adjustments would have little or no affect on correlations between TPT scores and other variables. The major reasons for this finding were: a given question counted as only 1/19th of a form's total score, a particular reader graded only 33% to 50% of a question's answers, and an applicant had the same reader for only two to four questions.

SEQUENCE EFFECTS

The two Trial Practices Tests were very different than any other test most applicants had taken in law school and totally different from previous bar examinations. Thus, it was quite possible that scores on the first form taken might be influenced by the applicants becoming accustomed to the testing format. In other words, scores might be affected by possible sequence or "warm-up" effects.

A sophisticated statistical procedure was needed to investigate this issue because the applicants who took the forms in the A-B sequence had higher average GBX scores than those who took the forms in the opposite order (see Table 2). The first step in this procedure involved using the total sample of 861 applicants to compute an equation that would predict an applicant's score on a form on the basis of that applicant's Phase I GBX score. The percentage of variance predicted by these equations was 37% for Form A and 32% for Form B (as compared to the 41% of the variance in Essay scores that is predicted by MBE scores).

The next steps consisted of computing each applicant's predicted score on each form, subtracting his/her actual score on a form from the predicted score, and then determining (by means of an analysis of variance) whether the average of these difference scores was higher or lower on the first form taken than on the second form taken. The results of this analysis indicated that there were no practically significant relationship between an applicant's total score on a form and whether that applicant took the form first or second (the form by test center interaction accounted for only 2% of the variance in difference scores).

Because there were no practically important sequence effects, all the remaining analyses in this report were conducted on the total sample of applicants (rather than separately by test center).

ITEM ANALYSES AND TEST RELIABILITY

An analysis was conducted on each form to determine (1) the degree to which scores on one question on a form correlated with the sum of the scores on the other 18 questions on that form (called corrected item-total r 's or correlations), (2) the reliability (internal consistency) of each form, and (3) the reliability of a form's Content and Strategy subscores. The data in Table 3 indicate there was generally a low positive correlation among the questions; i.e., applicants who received high scores on one question also tended to receive high scores on another question. Only 2 items on Form A and 7 on B had corrected item total r 's below .20 (and only 1 item, #3 on Form B, had a corrected r below .10). Despite the previously noted problems with Item #7 on Form B, this item's corrected correlations were only slightly lower than the corrected r 's on the other items.

The average reliability of the total score on a single, 90-minute form was .64. This is much higher than the reliability of the total score on two, 60-minute GBX essay questions. The reliability of the 3 hour total test (i.e., Form A + B) was .73 (as determined by the .58 correlation between forms). This is the same level of reliability that is obtained by 6 hours of GBX essay questions! The much higher reliability per hour of testing time with the Trial Practices Test than with the Essay was probably due to the TPT having 38 questions and all of them were in the same subject area.

Table 3

SUMMARY OF ITEM ANALYSIS RESULTS

Statistic	Form A			Form B		
	Content	Strategy	Total	Content	Strategy	Total
Number of items	12	8	19*	12	7	19
Mean inter-item correlation	.11	.10	.10	.08	.10	.08
Mean corrected item-total correlation	.25	.23	.25	.21	.21	.22
Reliability (alpha)	.60	.51	.66	.53	.45	.63

* The score on Item #12 on Form A was used in computing both subscores on this form.

FACTOR ANALYSES

Preliminary analyses indicated that the scores on a Content question did not generally correlate more highly with the scores on other Content questions than they did with the scores on the Strategy questions. And, the scores on a Strategy question did not generally correlate more highly with the scores on other Strategy questions than they did with the scores on the Content questions. These findings suggest that the two types of questions were not really measuring different skills or abilities.

This issue was investigated further by subjecting the data to a statistical procedure called factor analysis. This procedure examined each question's correlation with every other question to determine whether there were questions that generally correlated more highly with each other than they did with other questions on the test. This analysis revealed there were several clusters of questions on each form. These clusters were closely associated with the segments. Specifically, how well an applicant performed on one question was generally more highly correlated with how well he/she performed on the other questions on that segment than with questions associated with the preceding or subsequent segments. This pattern was probably due at least in part to one reader evaluating all the answers an applicant wrote to one segment before grading the next applicant's answers.

SUBSCORE ANALYSIS

The two forms of the Trial Practices Test yielded nine scores: Total scores on Form A and B, Content subscores on Forms A and B, Strategy subscores on Forms A and B, Content and Strategy subscores across forms, and the combined Total Score across forms.

The correlation between the Content subscore on Form A and the Content subscore on Form B (.44), and between the forms' two Strategy subscores (.41) were essentially the same as the correlation between the Content subscore on one form and the Strategy subscore on the other form (.39 and .43). Moreover, the correlation of the total Content score across the two forms with the total Strategy score across forms (.59) was almost identical to the correlation between the total scores on the two forms (.58).

The foregoing results confirmed the factor analysis finding that there was no meaningful distinction (in terms of applicant performance) between the Content and Strategy subscores. Total scores on the forms as well as total subscores across forms also had similar patterns of correlations with other variables. In other words, all the different scores on the Trial Practices Test appeared to be measuring the same common set of skills. Thus, only the combined total score across forms is discussed in the analyses presented in the next chapter.

Chapter 3

RESULTS

OVERVIEW

This chapter discusses the relationship between an applicant's score on the Trial Practices Test (TPT) and that applicant's GBX scores and background characteristics. There also is a discussion of applicant attitudes about the TPT and whether scores on it are affected by presenting the case and questions via a video versus a scripted mode. The chapter ends with a summary of the major findings and a discussion of their implications for the use of a TPT type test on a bar examination.

RELATIONSHIP WITH GBX SCORES

The correlation between TPT and individual GBX essay question scores ranged from .35 to .49 (the correlation with the GBX evidence question was .39).

Table 4 contains the observed and corrected correlations between GBX and TPT scores. The latter coefficients show the strength of the underlying relationship among the abilities assessed by each test. These data indicate the correlation between TPT and GBX scores is about as strong as the correlation between the essay and MBE (.66 observed and .81 corrected). In short, the abilities needed to earn a high score on the TPT are similar to but not exactly the same as those measured by the GBX.

Table 4

OBSERVED AND CORRECTED CORRELATIONS BETWEEN TPT AND GBX SCORES

<u>Type</u>	<u>Essay</u>	<u>MBE</u>	<u>Total</u>
Observed	.63	.61	.68
Corrected	.82	.76	.83

RELATIONSHIP WITH BACKGROUND CHARACTERISTICS

An analysis of the relationship between TPT scores and applicant background characteristics indicated these relationships paralleled those obtained with GBX scores. For instance, the correlation of TPT scores with applicant age, sex, and repeater status were: -.40, -.03, and -.41. The corresponding values for Total GBX score were: -.41, -.03, and -.52. The TPT did not widen or narrow the gap in performance level between racial groups that is already present on the bar examination (and in law school). There was a .33 correlation between "being Anglo" and scores on both measures. And, applicants who graduated from ABA approved schools also had higher scores on both the TPT and GBX than did applicants from California accredited or non-accredited schools.

RELATIONSHIP WITH CLINICAL EXPERIENCE AND SKILLS

As noted in Chapter 1, the applicants participating in this research completed a questionnaire that inquired about their clinical training and experience prior to July 1980. An analysis of these data indicated that applicants who had some clinical experience had statistically significantly higher TPT scores than those who did not have such experience. For example, applicants who had some paid or volunteer experience on actual cases had an 8 point higher average total TPT score (.61 standard deviation units) than applicants who did not have such experience. This difference appeared to be due to the experience factor rather than general academic ability because there was almost no difference between the two groups in average GBX scores. However, the experience factor explained only 7% of the variance in TPT scores (compared to the 46% explained by Total GBX) and the number of hours of experience was not related to TPT scores. In other words, a small portion of the difference between two applicants' TPT scores could be accounted for by differences among their respective amounts of clinical training and experience. It was not possible with the data available for this study to determine whether the quality (as distinct from just the amount) of clinical experience would have accounted for some of unexplained differences in TPT scores among applicants.

A group of 59 applicants who took the TPT also participated two weeks later in an Assessment Center (Klein, 1982). This participation involved two full days of testing on simulated cases under standardized conditions. On one day, an applicant served as counsel for the plaintiff in one case and on the other day, as counsel for the defendant in a different case. On both days, the participant took several oral and written tasks. There was a .61 correlation between total Assessment Center and TPT scores. This correlation rose to .87 after correcting for the unreliability in both measures. Thus, the underlying relationship between them was about the same as the underlying relationship between TPT and GBX scores.

APPLICANT EVALUATIONS OF TPT

About 97% of the applicants who participated in the TPT also completed a questionnaire that inquired about their attitudes toward the test. An analysis of these data revealed that most applicants rated the TPT as:

- o A less adequate measure of their "legal knowledge" than the GBX (ratings were on a 5 point scale from 1 = very poor to 5 = very good; means were 2.84, 3.43, and 3.21 for TPT, Essay, and MBE).
- o A better measure than the GBX of their "ability to perform as an attorney" (means on 5 point scale were 3.02, 2.76, and 2.39 for TPT, Essay, and MBE).
- o Having more realistic case situations than the GBX (means on 5 point scale were 3.54, 2.96, and 2.76 for TPT, Essay, and MBE).

In terms of the adequacy of time limits, the TPT, Essay, and MBE all received a rating of 1.7 (on a scale from 1 = less than enough to 3 = more than enough). There were no consistent relationship between ratings and performance. The 79% of the applicants who attended a commercial bar review course of at least 5 weeks duration had higher TPT and GBX scores than the 21% who did not attend (r 's were .20 and .30, respectively).

PRESENTATION MODE STUDY

An investigation was conducted to explore whether TPT scores would be affected if the TPT were presented in a scripted rather than in a video mode. This study involved the 2940 applicants who participated in another experiment that was conducted in conjunction with the July 1980 examination (Klein, 1981b). These applicants were randomly assigned to four groups. A given group received the background materials for either Form A or B, the script for three or four segments on their assigned form, and the 6 to 7 questions that were associated with these segments. They were then given 35 minutes to write their answers. Thus, unlike the video mode, the scripted mode permitted an applicant to read and review a segment's dialogue, but not to see or hear what transpired.

Before the grading process began, the answers to the questions written under the scripted mode were merged with those written under the video mode. Further, the same type of answer books were used for the two modes. Thus, the readers did not know under which mode a given answer was written.

Table 5 shows the average score on each question that was presented in both modes as well as the difference between modes. These data indicate that applicants who answered the questions under the video mode generally had higher scores than those who answered the same questions under the scripted mode. This finding was quite unexpected because all four groups taking the scripted mode had higher average GBX scores than the video group (by about 0.1 standard deviation units) and they could review the dialogue for a segment.

The relative advantage of the video mode tended to be greater on the higher numbered questions than on the lower numbered ones. This trend, which was most apparent in Groups 1 and 3, suggests the differences between modes may have been due to warm up effects (i.e., applicants taking the video had more opportunity to become accustomed to the novel question types). Other explanations for this pattern are that the scripted version allowed less time to answer per question than the video mode and/or that applicants in the scripted mode did not budget their time evenly across questions. It also is possible that knowledge of what transpired in the initial segments of a form helped the applicants in the video mode answer the questions on the subsequent segments on that form.

There was a .42 correlation between Total GBX scores and a scaled version of the scripted score (the scaling was done to control for some scripted groups having more questions than others and the average difficulty of the questions not being equal across groups). There was a .78 correlation between these measures after their scores were corrected for unreliability (where the correction involved estimating the reliability of a 6.5 item scripted test on the basis of the reliability of the 38 item total score).

The relatively strong corrected correlation between Total GBX and TPT scores regardless of presentation mode suggests there probably would be a fairly high correlation between scores on the two modes. However, because a given applicant only participated in one mode, there is no way of computing the actual correlation between modes.

Table 5

AVERAGE SCORE ON EACH QUESTION BY PRESENTATION MODE

Script Group	Test form	Question number	Video mode	Script mode	Difference
1	A	9	2.66	2.86	-.20
		10	2.01	2.18	-.17
		11	2.86	2.92	-.06
		12	2.66	2.48	.18
		13	2.57	2.17	.40
		17	<u>2.70</u>	<u>1.91</u>	<u>.79</u>
		Average	2.58	2.42	.16
2	A	7	2.66	2.85	-.19
		8	3.13	2.74	.39
		14	2.37	2.20	.17
		15	2.30	1.89	.41
		18	2.76	2.39	.37
		19	2.73	2.26	.47
		20	<u>3.39</u>	<u>2.02</u>	<u>1.37</u>
		Average	2.76	2.33	.43
3	B	5	3.15	3.46	-.31
		6	2.94	3.06	-.12
		7	2.86	2.70	.16
		8	3.10	2.93	.17
		11	2.87	2.42	.45
		12	<u>2.56</u>	<u>1.92</u>	<u>.64</u>
		Average	2.91	2.75	.16
4	B	4	1.87	1.84	.03
		9	2.58	2.66	-.08
		10	2.59	2.51	.08
		14	2.54	2.20	.34
		15	2.20	1.84	.36
		18	3.01	2.46	.55
		19	<u>2.57</u>	<u>2.18</u>	<u>.39</u>
Average	2.48	2.24	.24		
All	A+B	Average	2.68	2.44	.24

The standard deviation of a question under both modes was generally close to 1.00. There were about 735 applicants in each scripted group.

SUMMARY OF RESULTS AND CONCLUSIONS

Two forms of a Trial Practices Test (TPT) were developed in order to provide measures of certain important lawyering skills that were not tested (at least not directly) by California's General Bar Examination (GBX). The results of this research indicated:

- o TPT answers were scored just as reliably as Essay answers on the GBX. The scores on one half day of TPT tests were just as reliable as the scores on one full day of the GBX Essay.
- o How well an applicant performed on one form of the TPT was a good predictor of how well the applicant performed on the other form (observed and corrected correlations were .58 and .90). There were no practically important sequence effects.
- o Although the TPT included both Content and Strategy questions, there was no distinction between these question types in terms of applicant performance. An applicant's score on one type of question was just as highly correlated with scores on other questions of the same type as with scores on questions of a different type.
- o The TPT was just as highly correlated with the Essay and MBE as the Essay and MBE were correlated with each other. Thus, the TPT was measuring skills that were similar to but not exactly the same as those assessed by the GBX.
- o TPT and GBX scores had very similar correlations with various applicant background characteristics, such as race, sex, age, repeater status, and law school type. Thus, it did not widen or narrow the gap in average scores that currently exists among groups on the GBX.
- o Applicants who had some clinical experience did better on the TPT than those who had no experience. This difference was not attributable to differences in average GBX scores among experience groups. There also was a strong underlying relationship between TPT and Assessment Center scores.
- o Although applicants thought the GBX was a better test of their legal knowledge than the TPT, they had the opposite opinion regarding the relative efficacy of the measures as tests of an applicant's ability to perform as an attorney. They also felt the TPT presented more realistic case situations than the GBX.
- o Average scores on TPT answers were related to presentation mode (video versus scripted), but the nature of this relationship suggested it may have been produced by factors that were related to the experimental procedures rather than to the mode per se. Nevertheless, it appears the two modes are likely to yield similar results.

The foregoing findings suggest that including a TPT in the GBX would increase the GBX's reliability, make a small but noticeable change in who passed, and increase the validity of the examination by measuring skills that are relevant to practice but which are not now assessed.

APPENDIX A

EXCERPTS FROM NOTICE TO APPLICANTS

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 description (of the video problem):

Applicants in this session will take two problems. In each problem, applicants will be provided some written materials containing some relevant and some irrelevant factual and legal information about a case and will be allowed an appropriate period of time to review those materials. Thereafter, short segments of the examination of a witness or witnesses in a trial or arbitration proceedings will be shown on television monitors. After each segment, applicants will be asked one or more questions concerning such matters as objections and other procedures which may or may not have been utilized by counsel at that point or might be utilized later in the case and to justify their answers. Applicants also will be asked to evaluate the effectiveness of some of the devices and arguments employed by counsel depicted on the tape. Each problem will take one and one-half hours to complete (including the time to view the videotape).

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.

APPENDIX B

PREEXAMINATION QUESTIONNAIRE

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 research purposes. We are most appreciative of your cooperation.

1. Birthdate (/ /)

2. Which of the following jobs, if any, have you held for one or more months? Check all that apply:

- | | |
|--|---|
| <input type="checkbox"/> Attorney | <input type="checkbox"/> Legal Investigator |
| <input type="checkbox"/> Court Reporter | <input type="checkbox"/> Legal Secretary |
| <input type="checkbox"/> Law Clerk | <input type="checkbox"/> Paralegal assistant |
| <input type="checkbox"/> Law Enforcement Officer | <input type="checkbox"/> Other Law Related Employment |

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 | [5] None of these or no |
| [2] Japanese | [4] Taglog/Phillipino | 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 | [4] High School Graduate |
| [2] 6th to 8th | [5] College Graduate |
| [3] 9th to 11th | [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 | [4] High School Graduate |
| [2] 6th to 8th | [5] College Graduate |
| [3] 9th to 11th | [6] Graduate/Professional School Degree |

7. Circle the box below that best describes your undergraduate major:

- [1] Economics, Business, Accounting
- [2] Physical Science, Engineering, Mathematics, Biology
- [3] Social Science (Anthropology, Psychology, Sociology)
- [4] History, Government, Political Science
- [5] English, Journalism, Classical Studies, Philosophy
- [6] Fine Arts
- [7] Education
- [8] Other

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 more than 100 hours doing the activity.

	SIMULATED CASES	ACTUAL CASES Supervised Law School Programs	Paid or Volunteer Employment
a. Conduct legal research	[]	[]	[]
b. Prepare briefs, petitions, or motions	[]	[]	[]
c. Conduct direct examinations	[]	[]	[]
d. Conduct cross examinations	[]	[]	[]
e. Interview a client or 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
- [0] Other

11. Do you hereby apply to participate in the two day Assessment Center?

- (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 24 (inclusive) except for the dates circled below.

<u>Circle one</u>	Mon	Tue	Wed	Thu	Fri	Sat	Sun
in (1) Los Angeles only	4	5	6	7	8	9	10
in (2) San Francisco only	11	12	13	14	15	16	17
in (3) either LA or SF	18	19	20	21	22	23	24

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 Barelvas 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.

	Very Poor	Poor	Fair	Good	Very Good	Did Not Take
a. Special Session	1	2	3	4	5	9
b. Essay Examination	1	2	3	4	5	9
c. Multistate (MBE)	1	2	3	4	5	9

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.

	Very Poor	Poor	Fair	Good	Very Good	Did Not Take
a. Special Session	1	2	3	4	5	9
b. Essay Examination	1	2	3	4	5	9
c. Multistate (MBE)	1	2	3	4	5	9

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

	Less Than Enough	About Right	More Than Enough	Did Not Take
a. Special Session	1	2	3	9
b. Essay Examination	1	2	3	9
c. Multistate (MBE)	1	2	3	9

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."

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Did Not Take
a. Special Session	1	2	3	4	5	9
b. Essay Examination	1	2	3	4	5	9
c. Multistate (MBE)	1	2	3	4	5	9

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?

REFERENCES

- 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.
- Gulliksen, Harold Theory of Mental Tests. New York: Wiley, 1950.
- Klein, S. An analysis of grading practices on the California Bar Examination. Report prepared for the Committee of Bar Examiners of the State Bar of California, 1977.
- Klein, S. A comparison of the effectiveness of a single versus multiphased grading system. Report prepared for the Committee of Bar Examiners of the State Bar of California, 1980a.
- Klein, S. P. Intra- and inter-reader agreement on the essay section of the California State Bar Examination. Report prepared for the Committee of Bar Examiners of the State Bar of California, 1980b.
- Klein, S. P. Testing research skills on the California Bar Examination. Report prepared for the Committee of Bar Examiners of the State Bar of California and the National Conference of Bar Examiners, 1981a.
- Klein, S. P. 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 the National Conference of Bar Examiners, 1981b.
- Klein, S. P. An analysis of possible variations in pass/fail standards on the California State Bar Examination. Report prepared for the Committee of Bar Examiners of the State Bar of California, 1981c.
- Klein, S. P. and R. Bolus. An analysis of the relationship between clinical legal skills and bar examination results. Report prepared for the Committee of Bar Examiners of the State Bar of California and the National Conference of Bar Examiners, 1982.