



The State Bar of California

OPEN SESSION AGENDA ITEM 50-5 MARCH 2022

DATE: March 24, 2022

TO: Members, Board of Trustees

FROM: Amy C Nuñez, Program Director, Office of Admissions

SUBJECT: Approval of Guiding Principles in Examination Development to Minimize Potential Bias in Bar Exam Questions

EXECUTIVE SUMMARY

In 2019, the State Bar contracted with Scantron Corporation to conduct a differential item functioning (DIF) analysis for essay and performance test questions administered on the California Bar Exam between 2009 and 2019. The study raised no major areas of concern related to how the essays and performance tests performed in different populations. However, as some individual questions did result in some degree of differential performance, the report recommended that the State Bar conduct a bias and sensitivity review of those items to inform future item development. The report also recommended that the State Bar incorporate bias and sensitivity reviews into the examination review process prior to the administration of future exams.

To that end, the Board of Trustees directed State Bar staff to form a joint working group comprised of members from the Committee of Bar Examiners (CBE) and Council on Access and Fairness (COAF) to review the findings of the DIF study, conduct a bias and sensitivity review of flagged items, and develop a set of guiding principles to inform future question development. The DIF working group met from December 2020 through January 2022 and developed the attached Guiding Principles in Examination Development to Minimize Potential Bias in Bar Exam Questions, which were approved by the CBE in January 2022 and now come to the Board of Trustees in accordance with a Board resolution. The delivery of these guidelines aligns with the 2017–2022 Strategic Plan Goal 4, objective I, which states: “No later than March 31, 2020, identify ways that diversity and inclusion principles can be institutionalized in Bar exam development and grading analyses with final proposals, including any formal guidelines or rule

proposals, to be submitted to the Board by December 31, 2020.” The guidelines were delayed due to the numerous California bar exam-focused studies scheduled at that time, as well as the need to synthesize the findings from each of these initiatives. They included the California Attorney Practice Analysis, the DIF study, the review of the administration of the bar exam conducted by the Department of Consumer Affairs, and the review of the phased grading process. A review of all of these recommendations led to the establishment in May 2020 of the Blue Ribbon Commission and the joint committee working group focused on DIF.

BACKGROUND

The California bar exam, administered twice each year in February and July, consists of these parts:

- Essay questions
- Performance test
- Multistate Bar Examination – 200 multiple-choice items

Since 2009, the written portion of the bar exam has covered 13 subjects: Business Associations, Civil Procedure, Community Property, Constitutional Law, Contracts, Criminal Law and Procedure, Evidence, Professional Responsibility, Real Property, Remedies, Torts, Trusts, and Wills and Succession.

The State Bar contracted with Scantron Corporation to conduct a DIF analysis for essay and performance test items administered as part of the bar exam between 2009 and 2019—a total of 20 exams. DIF is a statistical procedure designed to evaluate performance patterns on specific questions across groups of test-takers. The purpose of a DIF analysis is to assess whether different groups with similar overall scores perform differently on test items.

When DIF is identified in a specific question, it denotes disparate performance within the test-taking population. Analyses such as these typically seek to assess impacts on performance by race and ethnicity. DIF is not synonymous with bias, and the presence of DIF in a test question does not necessarily mean a question is biased. No statistic can determine whether or not a question is biased. Ultimately, DIF is a means to help ensure the fairness of tests.

The purpose of the analysis was to examine DIF on the essay and performance test items associated with the following demographic variables: gender, race and ethnicity, and law school type. Secondary variables, such as item type, subject matter, and administration window were also considered in the analysis. First-time applicants who sat for the bar exam between July 2009 and February 2019 were included in the DIF analysis. A total of 116 essay and 36 performance test items were assessed in the study. The final report was presented to the CBE at its April 2020 meeting and to the Board at its May 2020 meeting.

The results of the DIF study indicated no major areas of concern for the bar exam with respect to DIF. Only 5 percent of items were flagged in the race and ethnicity analysis based on the Caucasian compared to non-Caucasian analysis, which is a small amount of DIF for an

assessment.¹ Although a large proportion of items were identified for law school-type DIF, those findings were not unexpected given that applicants at different school types have different educational backgrounds.

In May 2020, the Board directed the CBE to work with the COAF to convene a panel charged with reviewing those questions flagged for DIF in the 2020 DIF analysis and developing guidelines for minimizing the risk of future DIF.

DISCUSSION

The DIF working group began its work in December 2020 to review flagged items from the DIF analysis report and to develop guiding principles to mitigate against DIF on the bar exam prospectively. As part of this assignment, the DIF working group received an orientation about DIF analysis. Group members reviewed and rated essays and performance tests with high and low DIF ratings, both individually and collectively, in order to investigate factors that may be contributing to the differential performance. This work culminated in the attached Guiding Principles in Examination Development to Minimize Potential Bias in Bar Exam Questions.

The DIF working group held public meetings in 2020 (December 18), in 2021 (April 14, October 19, and November 29), and in 2022 (January 12). In between its April 2021 and October 2021 meetings, working group members reviewed specific bar exam questions that had been flagged for high and low DIF, with the purpose of identifying any bias that contributed to the differential performance.

The questions were evaluated based on the following dimensions for potential biases:

1. **Gender bias** – Does the exam question contain any language or concept suggesting any gender bias?
2. **Cultural bias** – Does the exam question contain any language or concept suggesting any cultural bias, with “culture” broadly defined?
3. **Offensive** – Does the exam question contain any language or concept that may be found offensive for any reason?
4. **Perpetuating stereotypes** – Does the exam question contain any language or concept that may be perceived as perpetuating certain stereotypes?

¹ A number of DIF analyses were conducted with respect to race and ethnicity variables. In all cases, Caucasian was designated as the reference variable and the focal groups for each analysis were Black/African American, Asian, Hispanic, or Other. However, the difference between the number of candidates in the reference was significantly larger than the focal groups; therefore, a variable more comparable in sample size to the Caucasian reference group was created by combining the Black/African American, Asian, Hispanic, and Other categories into a single variable.

5. **Unfair representation of certain groups** – Does the exam question contain any language or concept that appears to represent certain groups unfairly?
6. **Construct irrelevant knowledge required** – Does the exam question, in the language used or how the issues are presented, appear to require applicants to be knowledgeable in certain areas that are irrelevant for the main issues in the question to be answered with competence?
7. **Language that put any group at an advantage** – Does the exam question contain any language that may put certain groups at an advantage or a disadvantage?

The DIF working group member review results were compared with the actual DIF study findings. Results of the working group exercise showed that the frequency of the various dimensions flagged by group members were not significantly different between the low and high groups. Overall, the assessment results from the group members did not reflect the results of the DIF study. However, the members identified areas where the questions could improve on clarity, reducing irrelevant information, and where exam development should strive to use concepts that are more universally understood. Most of the language in individual questions flagged by working group members as raising concern was relatively subtle. No racially inappropriate, gender inappropriate, or otherwise overtly biased terminology was found. Nor did the exercise reveal overt stereotyping or favoring one racial category over another.

The key themes identified were refined into specific principles during subsequent working group meetings, which eventually resulted in the Guiding Principles in Examination Development to Minimize Potential Bias in Bar Exam Questions that are being presented to the Board for approval in this agenda item (see Attachment A).

The guiding principles include an important overarching statement of the goal for test questions and the test as a whole:

Examination developers should strive to make tests that are as fair and accessible as possible for all groups of test-takers, despite differences in characteristics that include, but are not limited to, ethnic group, gender or gender identity or expression, sexual orientation, regional background, native language, race, religion, age or generational differences, socioeconomic status, disability status, and veteran status. The goal of fairness in testing is to assure that test materials are as free as possible from bias and do not impose unnecessary barriers to the success of diverse groups of test-takers. To test knowledge of the law and the ability to apply the law, test questions should be designed to ensure that the nonlegal aspects of any question are broadly understandable by the population of test-takers and do not contain information that will have the effect of disadvantaging one population of test-takers vis-à-vis the others.

The document then sets forth 14 guiding principles to minimize bias in questions included on the bar exam. The first 10 are intended to serve as guidance in question development and will be provided to all developers of the bar exam—including question drafters and pre-testers, the Examination Development and Grading Team, and the Performance Test Drafting Team. The guidelines refer to the content of exam questions, and what should be done and what should be avoided to minimize potential bias. As such, the guidelines will also be part of what CBE considers during the question selection process. Examples of guidance provided by these principles include: using scenarios that are accessible or equally familiar to all; using terminology that has common meaning and understanding; and ensuring balanced representation of diverse and nondiverse groups in situations where individuals commit misconduct or are on the receiving end of such conduct.

The last four guiding principles are intended to help provide the background and training that are beneficial to being able to carry out the previous guiding principles. These principles describe annual training that exam developers should receive; the goals for question-developer recruitment practices; and recommendations to the CBE for further review and development of its question-developer and pre-tester recruitment processes. The overall objective is to advance diversity and inclusion in the pool to ensure varied perspectives, backgrounds, and experiences, while also maintaining the necessary level of examination development, grading, and pretesting skills and expertise.

FISCAL/PERSONNEL IMPACT

None

AMENDMENTS TO RULES

None

AMENDMENTS TO BOARD OF TRUSTEES POLICY MANUAL

None

STRATEGIC PLAN GOALS & OBJECTIVES

Goal: 4. Support access to legal services for low- and moderate-income Californians and promote policies and programs to eliminate bias and promote an inclusive environment in the legal system and for the public it serves, and strive to achieve a statewide attorney population that reflects the rich demographics of the state's population.

Objective i: No later than March 31, 2020, identify ways that diversity and inclusion principles can be institutionalized in bar exam development and grading analyses with final proposals, including any formal guidelines or rule proposals, to be submitted to the Board by December 31, 2020.

RECOMMENDATIONS

None

ATTACHMENT LIST

- A.** Guiding Principles in Examination Development to Minimize Potential Bias in Bar Exam Questions

GUIDING PRINCIPLES IN EXAMINATION DEVELOPMENT TO MINIMIZE POTENTIAL BIAS IN BAR EXAM QUESTIONS

Examination developers should strive to make tests that are as fair and accessible as possible for all groups of test takers, despite differences in characteristics that include, but are not limited to, ethnic group, gender or gender identity/expression, sexual orientation, regional background, native language, race, religion, age or generational differences, socioeconomic status, disability status, and veteran status. The goal of fairness in testing is to assure that test materials are as free as possible from bias and do not impose unnecessary barriers to the success of diverse groups of test takers. To test knowledge of the law and the ability to apply the law, test questions should be designed to ensure that the nonlegal aspects of any question are broadly understandable by the population of test takers and do not contain information that will have the effect of disadvantaging one population of test takers vis-à-vis the others.

To that end, the following guiding principles to enhance fairness, provide equity in accessibility, and minimize potential bias in the development of questions on the California Bar Examination should be followed:

- Describe contexts and scenarios that are accessible or equally familiar to all and use terminology that has common meaning and understanding; provide short explanations or definitions for acronyms and uncommon terms when use of a commonly understood term is not feasible.
- Ensure balanced representation of diverse and nondiverse groups in both positive and negative situations, in situations where individuals commit misconduct or are on the receiving end of misconduct, or where individuals are perpetrators and victims or witnesses.
- Ensure gender balance in the fact patterns and use gender-neutral names, where appropriate.
- Ensure balanced or neutral representation across gender, culture, ethnic groups, and disability status.
- Avoid using cultural references or idiomatic expressions that may not be equally familiar to all test takers.

- Avoid construct-irrelevant content.
- Avoid content that unnecessarily angers, offends, upsets, or otherwise distracts test takers; in that regard, minimize the use of fact patterns that involve unnecessarily sensitive or emotionally-charged topics.
- Avoid use of offensive language and stereotypes.
- Avoid language in questions that assumes all test takers have the same life experiences, socioeconomic status, or family background.
- In considering fairness and bias, sets of test questions should be reviewed as a group for each exam and names and scenarios adjusted to ensure a balanced exam.
- All developers of the California Bar Examination should participate in or receive specialized training each year, which may include subjects such as implicit bias and stereotype threat or any other subjects that are relevant to recognizing and addressing potential bias.
- The recruitment practices for all question developers should endeavor to create a pipeline of potential developers with varied perspectives so as to advance diversity and inclusivity, while also maintaining the necessary level of examination development and/or grading skills and expertise.
- Recommend that the Committee of Bar Examiners review all exam developer and grader recruitment processes and revisit term limits so as to advance diversity and inclusivity in the pool of question developers to include varied perspectives, backgrounds and experiences, while also maintaining the necessary level of examination development and/or grading skills and expertise.
- Recommend that the Committee of Bar Examiners develop recruitment processes for exam question pretesters so as to advance diversity and inclusivity in the pool of pretesters to include varied perspectives, backgrounds and experiences, while also maintaining the necessary level of pretesting skills and expertise.