



The State Bar *of California*

**OPEN SESSION
AGENDA ITEM O-202
JANUARY 2022
COMMITTEE OF BAR EXAMINERS III.D**

DATE: January 28, 2022

TO: Members, Committee of Bar Examiners

FROM: Lisa J. Cummins, Principal Program Analyst, Examination Development

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 (CBX) between 2009 and 2019. The resulting report concluded that the study raised no major areas of concern for the CBX with respect to how the essays and performance test performed in different subgroups (hereafter referred to as DIF). However, since 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 identified as having some differential performance to inform future item development. The report also recommended that the State Bar to incorporate bias and sensitivity reviews into the examination review process prior to the administration of future exams.

To that end, the Board of Trustees (Board) 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, to conduct a bias and sensitivity review of flagged items, and to develop a set of guiding principles to inform future CBX 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 are being presented to the CBE for approval.

BACKGROUND

The California Bar Examination (CBX), administered twice each year in February and July, consists of the following parts:

- Essay Questions
- Performance Test
- Multistate Bar Examination (MBE) – 200 multiple-choice items

Since 2009, the written portion of the CBX has covered the following 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 CBX between 2009 and 2019 (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, the information helps to identify potential issues in the question. 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, it is a means to help ensure the fairness of tests.

The purpose of the analysis was to examine DIF on the CBX essay and performance test items with respect to the following demographic variables: gender, race/ethnicity, and law school type. Secondary variables, such as item type, subject matter, and administration window were also considered in the analysis. The report was presented to the CBE at its April 2020 meeting and to the Board at its May 2020 meeting.

First-time applicants who were administered the CBX between July 2009 and February 2019 were included in the DIF analysis. A total of 116 essay and 36 performance test items were included in the study.

The demographic variables included in the DIF analysis were gender (Male, Female), race/ethnicity (Asian, Black/African American, Hispanic, Caucasian, Other), and law school type (California ABA, California Accredited, California Registered, Out-of-State ABA, Attorney, Foreign JD).

The results of the DIF study indicated no major areas of concern for the CBX with respect to DIF. Only 5% of items were flagged in the race/ethnicity analysis based on the Caucasian/Non-Caucasian analysis, which is a small amount of DIF for an assessment. 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 Committee of Bar Examiners to work with the Council on Access and Fairness to convene a panel charged with reviewing questions flagged for DIF in the 2020 differential item function analysis and developing guidelines for minimizing the risk of future differential item functioning.

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 CBX prospectively. As part of this assignment, the DIF working group received an orientation about DIF analysis; they 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, November 29), and in 2022 (January 12). In between its April 2021 and October 2021 meetings, individual working group members were given an exercise to review the same 20 past bar exam questions (18 essay questions and 2 PTs) selected randomly from the DIF study by State Bar Office of Research and Institutional Accountability (ORIA) staff as having highest and lowest DIF flags (10 each). Subgroup members were asked to analyze all 20 questions for potential biases in the seven dimensions described below, and thereafter to meet with staff to discuss their findings. Results were received from 7 members, generating 980 (20 x 7 x 7) data points to learn from the exercise.

The 20 questions were evaluated in 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?
- 5) **Unfair representation of certain groups** – does the exam question contain any language or concept that appear to represent certain groups unfairly?

- 6) **Construct irrelevant knowledge required** – does the exam question, in the language used or how the issues are presented, appears 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 group at an advantage or disadvantage?

Ron Pi from ORIA presented the results of the subgroup exercise at the Working Group's October 19, 2021 meeting. Summed across all questions and Working Group members, gender was flagged most frequently, followed closely behind by stereotype. Three dimensions (offensive, representation, and language) were flagged at a significantly lower rate.

The DIF Working Group member review results were compared with the actual DIF study findings. The 20 exam questions that were used in the Working Group exercise had been grouped into the low vs. high category according to the frequency of DIF flags in the study. Working Group members were not told which questions had the high flags and which had the low flags. Interestingly, ORIA's analysis of the results of the Working Group exercise showed that the frequency of the various dimensions flagged by Working Group members were not significantly different between the low and high groups. Overall, the assessment results from the Working Group members did not reflect the results of the DIF study and the results of the DIF study indicated no major areas of concern for the CBX with respect to DIF.

Most of the language in individual questions flagged by Working Group members as raising concern were relatively subtle. No racially inappropriate or otherwise overtly biased terminology was found.

During the Working Group subgroup and full group discussions, staff took notes to summarize the findings and to develop themes. The key themes were fleshed out and 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 for CBE approval in this agenda item.

The first ten bullet points in this document are to serve as guidance in CBX question development and will be provided to all developers of the CBX (including question drafters and pretesters, the Examination Development and Grading (EDG) Team, and the Performance Test (PT) 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 CBX considers during the question selection process.

The last four bullet points describe the annual training exam developers should receive, the goals for question developer recruitment practices, and recommendations to the CBE for further review/development of its question developer and pretester recruitment processes so as 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

RECOMMENDATION

It is recommended that the “Guiding Principles in Examination Development to Minimize Potential Bias in Bar Exam Questions” developed by the DIF Working Group be approved, then sent to the Board of Trustees with a recommendation for approval.

PROPOSED MOTION

Should the Committee of Bar Examiners agree with the staff recommendation, the following motion would be appropriate:

MOVE that the “Guiding Principles in Examination Development to Minimize Potential Bias in Bar Exam Questions” developed by the DIF Working Group be approved, and submitted to the Board of Trustees with a recommendation for approval.

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.