

4.2 Discussion of Bar Exam Multiple-Choice Question Development Process



The State Bar of California

Discussion of Bar Exam Multiple-Choice Question Development Process

Committee of Bar Examiners Meeting April 18, 2025

Presented By:
Donna Hershkowitz, Chief of Admissions

Question Development & Content Validation Overview

Content Validation Panel

- Panelists individually review MCQs for:
 - Content Accuracy
 - Cognitive Complexity
 - Potential Bias
 - Minimum Competency Alignment
- Style edits applied to all questions
- Group discussion to achieve consensus

Review + Edit

- Report produced by psychometric facilitator with panel feedback
- Final recs cited as Ok, Revised, Replace
- Feedback exchange with subject matter expert

Item Bank/ Question Selection

- Items reviewed for grammar, style, and revision accuracy during finalization
- Percentage allocation breakdowns met during question selection

Post-Exam Analysis

- Conducted with psychometricians
- Deep review of item performance

Going Forward:

Professional copyediting to address spelling errors and typos

Adding additional layer of review for content accuracy

MCQ Performance Overview



Reliability Findings

- Target Reliability: 0.80+
- Actual (171 scored items): 0.89
- Initial (Nov MCQs): 0.82



Item Difficulty

- The proportion of applicants that answered a given question correctly



Item Discrimination

The ability of an item to differentiate among test takers on the basis of how well they know the material being tested

Response Option Analysis

- Examination of whether the answer choices (options) functioned properly - were the distractors reasonable

Calculations by Subject Area

	Target	Civil Procedure	Con Law	Contracts	Criminal Law & Procedure	Evidence	Real Property	Torts	Total
Average Difficulty	0.30 to 0.80	0.60	0.64	0.64	0.59	0.68	0.60	0.63	0.63
Average Discrimination	0.10+	0.19	0.22	0.19	0.13	0.20	0.18	0.13	0.18
Performance Flags*	< 6 per subject area	3	5	4	9	2	6	11	40

*Items were flagged for performance if they had item difficulty values outside the target range, average item discrimination below the target value, and/or fewer than three functioning response options defined as having at least 5% of applicants respond to the option (i.e., A, B, C, or D).

Questions?