

Psychometric Analysis – GBX

*February 2025 GBX Preliminary
Results*





General Bar Exam (GBX)

Steps in development and validation



GBX:

- Confirm exam blueprint (complete)
- Develop & review new questions (continuous)
- Pilot test sample of questions (complete)
- Construct exam forms (complete)
- Administer February 2025 bar examination (in progress)
- Analyze applicant data (in progress)
- Determine raw passing score (in progress)

Development & review of new questions

- Draft new questions relative to exam blueprint
- Editing for style and structure
- Qualified, external reviewers not affiliated with Kaplan
- Independent review of draft questions for:
 - Content accuracy
 - Cognitive complexity
 - Bias, diversity, and inclusiveness
 - Appropriate level (i.e., Minimally Competent Applicant)

Testing Industry Guidelines for Disruptions

- Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014)
 - Validity, reliability, and fairness of scores and decisions
- Technology-Based Assessment Guidelines (ITC & ATP, 2022)
 - Addressing test disruptions (Guidelines 3.33 – 3.44)
 - Responding to incomplete data that result from technology disruptions (Guidelines 4.20 – 4.24)

February 2025 GBX, Multiple Choice Question Analysis – Summary

- The February bar exam, multiple-choice questions, were administered in-person and online to 3,855 (2-day) applicants
 - 117 handwriting applicants
 - 3,738 applicants tested online (small test center, larger pop-up center, home, other)
- Multiple choice questions evaluated for:
 - Item difficulty – goal of values of 0.30 to 0.80
 - Item discrimination – goal of positive values of 0.10 and higher
 - Option analysis – goal of each distractor being plausible (0.05 or higher)
 - Projected internal consistency reliability – goal of 0.80 and higher

February 2025 GBX, Multiple Choice Question Analysis – Summary (cont.)

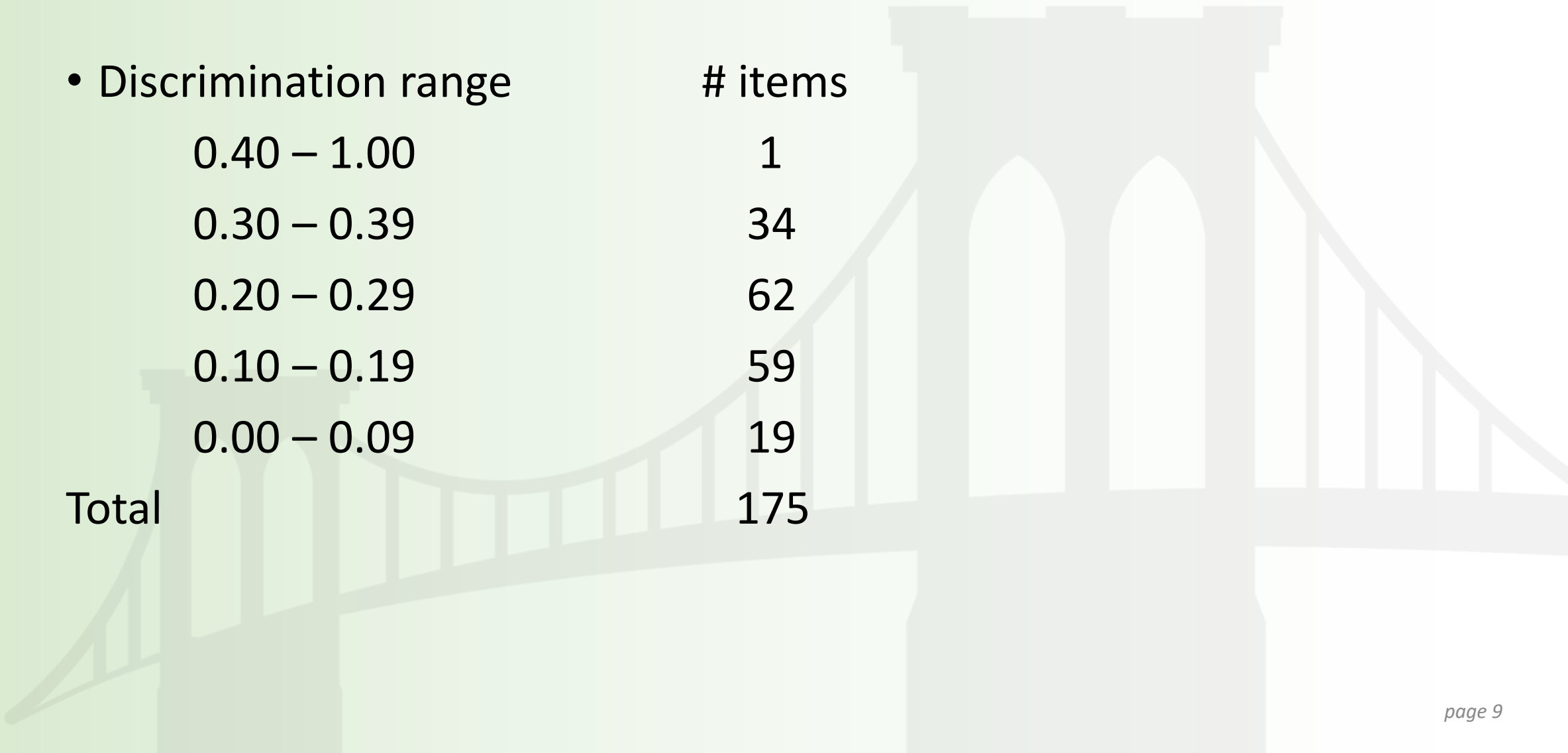
- Related comments
 - Responses for one question (KES00064) were not recorded for online administration applicants and removed from scoring.
 - Other questions were flagged for revision or removal.
 - 35 questions administered on the November experimental exam were embedded among the 200 MCQs administered in February to evaluate item drift
 - 175 of the 200 administered items were selected for scoring based on:
 - Subject area representation (25 questions each)
 - Acceptable statistical performance

Item difficulty summary (0.30 – 0.80)



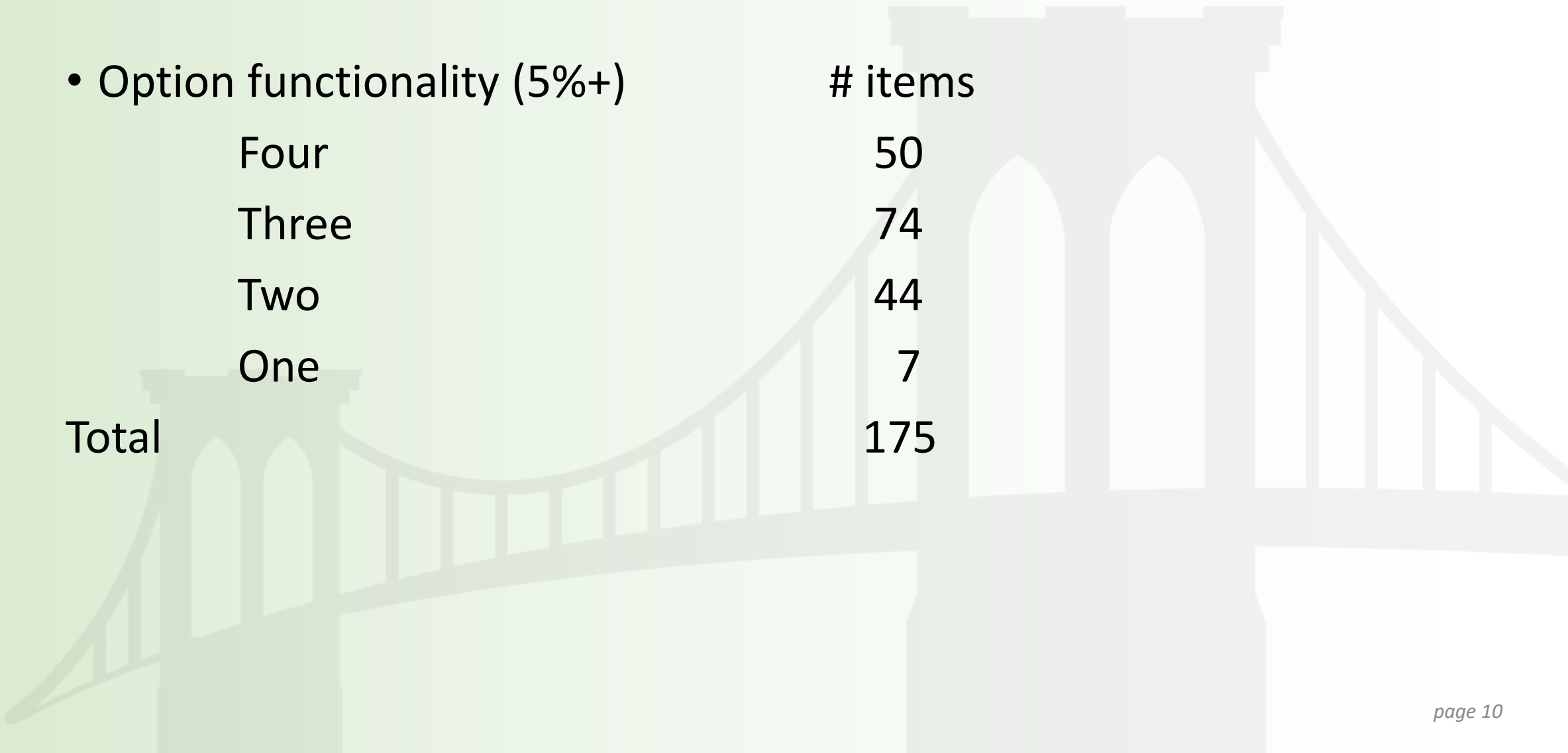
• Difficulty range	# items
0.81 – 1.00	33
0.70 – 0.80	47
0.60 – 0.69	39
0.50 – 0.59	23
0.40 – 0.49	12
0.30 – 0.39	14
0.00 – 0.29	7
Total	175

Item discrimination summary (0.10 and higher)



• Discrimination range	# items
0.40 – 1.00	1
0.30 – 0.39	34
0.20 – 0.29	62
0.10 – 0.19	59
0.00 – 0.09	19
Total	175

Response option analysis summary (3 to 4)



• Option functionality (5%+)	# items
Four	50
Three	74
Two	44
One	7
Total	175

Internal Consistency Reliability

- Coefficient alpha
 - Estimates the average split-half reliability of all possible combinations
 - Goal is 0.80+ for supporting individual decisions
 - Predicted reliability based on the November experiment was **0.82**
 - Actual reliability for the 175 scored items was **0.90**

Internal Consistency Reliability – Predicted

- Spearman-Brown Prophecy Formula
 - If the exam were reduced to 140 scored questions (20 for each subject area), the predicted reliability is **0.87**.
 - If exam were reduced to 105 scored questions (15 for each subject area), the predicted reliability is **0.84**.
- Future options for multiple-choice section
 - Multiple linear forms
 - Linear on the Fly (LOFT)
 - Adaptive

Responding to Test Disruption – Historical Practice

July 2021 California Bar Exam

- Pro rata grading adjustments
 - Baseline expectations determined from unaffected responses
 - Affected responses were adjusted up based on the estimated performance
- Regression
 - Applied to extensive disruptions (3 or more affected written sessions; 2 or more affected multiple-choice sessions)
 - Predicted performance was estimated for missing applicant data based on the estimated relationships among sessions and sections of the test

Responding to Test Disruption – Options

February 2025 California Bar Exam

- Differential pro rata grading adjustments
 - Baseline expectations determined from unaffected responses
 - Affected responses differentially adjusted based on the type and severity of the disruption
- Imputation method like regression
 - Applied to extensive disruptions (3 or more affected written sessions; 2 or more affected multiple-choice sessions)
 - Predicted performance was estimated for missing applicant data based on the estimated relationships among sessions and sections of the test
- Makeup score replacement methods
 - Makeup administration opportunity for applicants to replace missing scores