

Test Development & Validation – GBX

Managing change



General Bar Exam (GBX)



Steps in development and validation



GBX:

- Confirm exam blueprint
- Review existing questions
- Develop & review new questions*
- Construct exam forms*
- Collect applicant data
- Analyze data*
- Determine raw passing score*

Influencing factors

- Administration mode(s)
 - Event based, remote proctored, testing centers, combination
- Capacity to maintain administration timing
 - Same days, testing window, combination
- Number of applicants testing concurrently
 - Limits on administration modes
- Security considerations
 - Question bank size, number of forms, question exposure

Develop & review new questions

- Draft new questions relative to exam blueprint
- Review questions for:
 - Content and cognitive complexity
 - Bias, diversity, and inclusiveness
 - Appropriate level (i.e., minimally competent/qualified applicant)
- Construct forms with sets of common questions and pretest questions
 - Common question support equating across forms
 - Pretest questions are used to collect data to increase size of question bank

Analyze data

- Initial administration to collect baseline data (“operational pretest”)
- Multiple choice questions evaluated using:
 - Item response theory (IRT) – item parameters that are sample independent
 - Classical test theory (CTT) – item difficulty, item discrimination, option analysis
- Essay and PT questions evaluated using:
 - Intra- and Inter-Rater agreement indices
 - Embedded performances (validity papers)
- Evaluate technical quality to remove questions that do not function appropriately prior to finalizing scores
- Additional analyses:
 - Differential item functioning (DIF) to empirically evaluate potential bias
 - Mode effects to evaluate potential differences in delivery approaches
 - Forensic analysis (e.g., item drift, similarity analysis)

Raw scores and Scale scores



- Raw scores – number of points earned on a question or form of an examination
 - Associated with specific questions and forms of an examination
 - Not directly comparable unless the same questions are used every time
- Scale scores – transformation of raw scores to an interpretative scale
 - Intended to maintain a common interpretation of scores across forms of an examination
 - Fixed location on the scale is used as an anchor point for the passing score to maintain fairness through statistical equating

Options for GBX standard setting

- Key changes – replacement of multiple-choice items on the exam, lack of anchor items for equating the initial examination
- Could consider equipercentile equating to link to current examination
 - Not recommended because of normative interpretation
- Options to establish baseline raw passing score (no change to scale score)
 - Conduct study prior to administration
 - Content anchored but not informed by applicant data
 - Conduct study following administration
 - Content anchored and informed by applicant data
 - Initial judgments before and evaluation once student data area available
 - Content anchored, informed by applicant data, and reduces reporting time