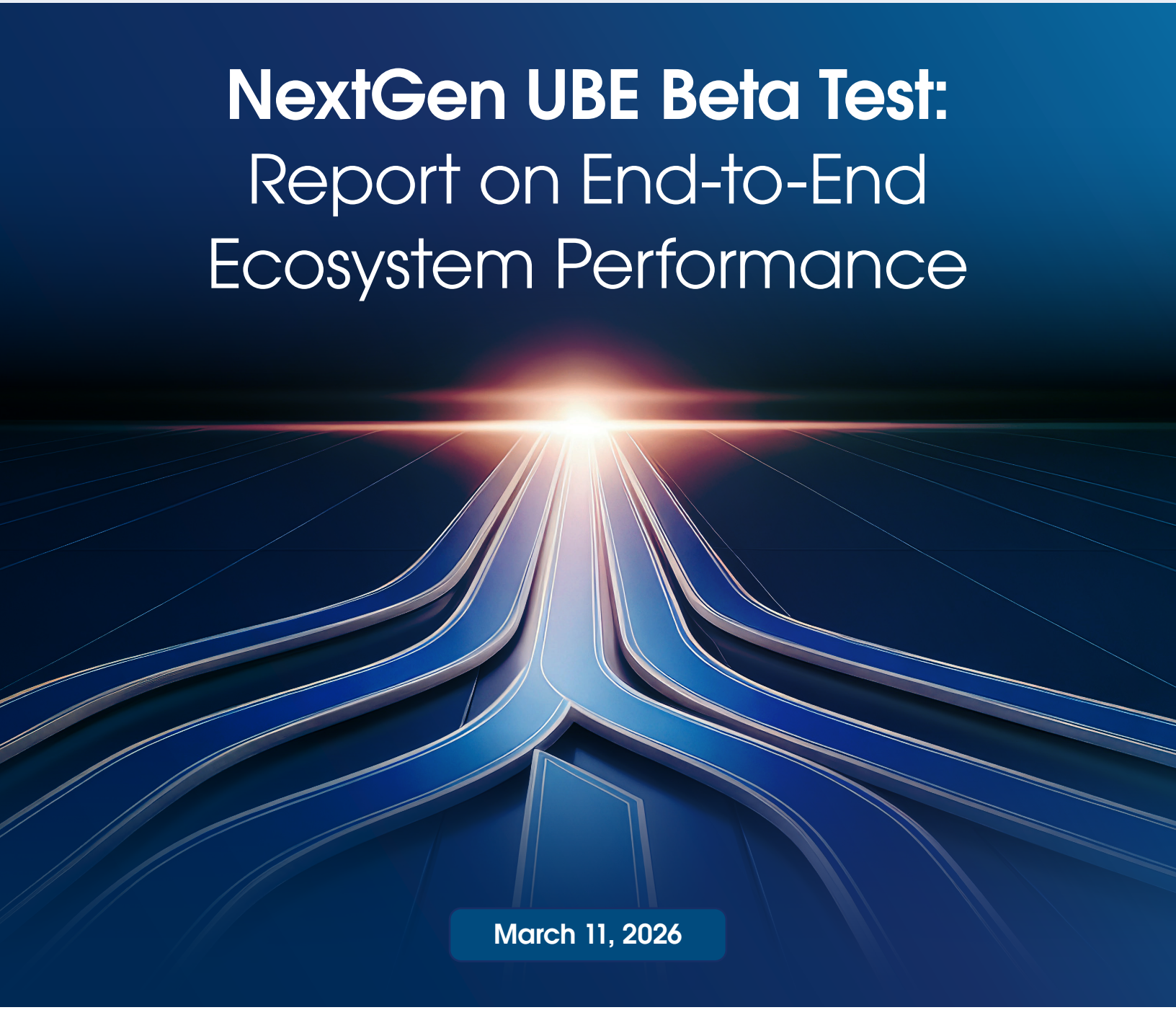
NextGen UBE™ Uniform Bar Examination

NextGen UBE Beta Test: Report on End-to-End Ecosystem Performance



March 11, 2026

NCBE® National Conference of Bar Examiners

The National Conference of Bar Examiners, founded in 1931, is a not-for-profit corporation that develops licensing tests for bar admission and provides character and fitness investigation services. NCBE also provides testing, research, and educational services to jurisdictions; provides services to bar applicants on behalf of jurisdictions; and acts as a national clearinghouse for information about the bar examination and bar admissions.

Our Mission

NCBE promotes fairness, integrity, and best practices in admission to the legal profession for the benefit and protection of the public. We serve admission authorities, courts, the legal education community, and candidates by providing high-quality

- assessment products, services, and research;
- character investigations; and
- informational and educational resources and programs.

Our Vision

A competent, ethical, and diverse legal profession.

NCBE
National Conference of Bar Examiners

Building a competent, ethical, and diverse legal profession.

NextGen UBE
Uniform Bar Examination

302 South Bedford Street
Madison, WI 53703

ncbex.org
thebarexaminer.org
ncbex.org/exams/nextgen
nextgenbarexam.ncbex.org

/ncbexaminers
company/ncbex
/ncbexaminers

Contents

Contents continued

Executive Summary

The January 2026 administration of the NextGen Uniform Bar Examination (NextGen UBE) beta test represented the first large-scale, live administration of the full NextGen exam ecosystem in conditions designed to reflect full operational rollout. The beta test evaluated the integrated delivery of exam content, digital platforms, jurisdiction administration workflows, candidate readiness processes, constructed-response grading operations, and end-to-end data capture and reporting.

A total of 1,512 examinees began testing in five locations across four jurisdictions, with 1,500 examinees (99.2%) completing all exam sections. Examinees were assigned to one of three complete exam forms constructed in accordance with the operational blueprint. The administration was observed and supported by 64 jurisdiction representatives from 40 jurisdictions, providing direct evaluation of administration workflows, platform functionality, and operational procedures.

Across the full exam lifecycle—including candidate readiness, exam delivery, response capture, grading workflows, and score reporting—core systems and operational processes functioned as designed. The Jurisdiction Portal, Candidate Portal, ITS Exam Day Portal, and secure delivery platform supported coordinated administration, real-time monitoring, and successful capture and transmission of examinee responses. Backup device workflows,

JANUARY 2026

NextGen UBE Beta Test

1,512 examinees, 99.2% completing all exam sections

Five locations across four jurisdictions

Observed by 64 representatives from 40 jurisdictions

All systems evaluated across the full exam lifecycle

144 graders scored more than 64,000 responses

Examinees found testing platform easy to use, exam content practical and relevant

Jurisdiction staff reported well-structured workflows and effective systems and policies

response preservation mechanisms, and escalation protocols functioned effectively when isolated disruptions occurred, allowing examinees to continue testing without loss of responses.

Constructed-response grading was conducted at operational scale, with 144 graders organized into 54 teams scoring more than 64,000 responses across 63 constructed-response questions. Grader training, calibration, assignment, reconciliation, and scoring workflows were successfully implemented using the digital grading platform, allowing evaluation of scoring operations under conditions approximating operational administration.

Examinees reported that the exam platform felt intuitive and easy to use and described the assessment as practical, professionally relevant, and reflective of legal reasoning and task-based judgment. Jurisdiction staff reported that administration workflows were well structured, monitoring systems provided effective real-time visibility, and exam-day policies were implementable under live conditions. Compared to the prototype administration, jurisdiction staff observed improved platform stability, clearer workflows, and more predictable administration processes.

Consistent with the purpose of a beta administration, the process also identified targeted opportunities for refinement, including enhancements to administrative reporting functionality, workflow visibility, user interface efficiency, and training and guidance materials. These refinements reflect normal product maturation following large-scale live testing and are focused on improving operational efficiency and administrative consistency.

This descriptive report documents operational delivery, system performance, and stakeholder experience during the beta administration. Detailed psychometric analyses—including reliability, dimensionality, subgroup analyses, constructed-response scoring statistics, and scaling and equating outcomes—will be presented in a separate technical report scheduled for release in May 2026.

Taken together, the findings indicate that the NextGen UBE ecosystem functioned cohesively under live administration conditions and support continued progression toward operational launch in July 2026, with final work focused on refinement and optimization of an operationally validated system.

Introduction

The January 2026 administration of the NextGen UBE beta test represents a pivotal milestone in NCBE's multi-year effort to modernize the bar examination while preserving its core purpose: protecting the public through a valid and reliable assessment of minimum competence. The full NextGen ecosystem was tested under live, large-scale, multi-site conditions that closely approximated an operational (live) administration.

Relationship Between the October 2024 Prototype and the January 2026 Beta

NCBE's development and readiness strategy for the NextGen UBE has been intentionally staged. Rather than rely on a single high-stakes transition, NCBE has built evidence and operational confidence through sequential large-scale administrations designed to answer different questions at different points in the testing arc. While NCBE had previously tested NextGen question types and content in shorter pilot and field tests, the October 2024 prototype exam and the January 2026 beta test were full-length exams designed to mimic operational test administrations. However, the two exams differed both in what they evaluated and in their purpose.

Purpose of the October 2024 Prototype: Validate the Exam Form and Integrated Assessment Model

The October 2024 prototype exam was designed primarily to evaluate the exam itself under live conditions. It provided NCBE with an opportunity to administer a full NextGen form at scale, collect performance data across item types, and evaluate whether the integrated assessment model—selected-response questions combined with scenario-based tasks and constructed responses—functioned as intended in a realistic testing context. The prototype also allowed early

evaluation of usability and content navigation features that are central to performance task completion and to ensuring the validity of a digital licensing exam.

In short, the October 2024 prototype answered the question: Does the NextGen exam design work as an assessment under live conditions?

Purpose of the January 2026 Beta: Validate the Full Operational Ecosystem

The January 2026 beta test was designed to evaluate something different: the integrated delivery of the full NextGen ecosystem under operationally representative conditions. Unlike the prototype, the beta intentionally exercised end-to-end operations across the full exam lifecycle, including candidate readiness workflows, multi-site exam delivery, real-time monitoring, response capture and transmission, constructed-response grading at scale, and the data pipelines and reporting needed to support jurisdiction use.

The beta was also designed to confirm operational resilience. Systems were tested under realistic load, contingency procedures were exercised, and escalation workflows were evaluated to ensure examinees could continue testing and responses could be preserved even when isolated disruptions occurred.

In short, the January 2026 beta answered the question: Can the full NextGen ecosystem be used reliably and consistently at operational scale?

The Progression from Prototype to Beta

The two administrations served complementary purposes. A full-form prototype can demonstrate that the assessment model is viable and produce critical performance data, but it does not fully test the operational machinery required for an exam that must function consistently across jurisdictions, sites, and user roles. Conversely, an operational beta can validate readiness workflows, monitoring systems, and grading operations, but it is most valuable once the exam form and interface design have been sufficiently proven through earlier large-scale work. Together, the prototype and beta represent a deliberate progression from assessment validation to operational validation.

The Beta Test

This beta followed several years of structured research, development, pilot and field testing, large-scale prototype administration, standard-setting and concordance preparation, and sustained collaboration with jurisdictions and subject-matter experts. Each phase of this testing arc was intentionally designed to answer defined research and operational questions, reduce uncertainty, and build evidence incrementally rather than rely on a single high-stakes transition.

The beta functioned as the bridge between development and operations. It integrated jurisdiction administration workflows, candidate readiness processes, finalized exam content, the launch-version digital delivery platform, administration policies, accommodations delivery, constructed-response grading at scale, and end-to-end data pipelines into a coordinated live administration. For many jurisdictions, it also provided the first opportunity to observe the complete lifecycle of the NextGen UBE—from roster creation through response capture, grading, and score reporting.

The entire beta cycle spanned five months, beginning with the opening of the Jurisdiction Portal on November 3, 2025, and concluding with beta score reporting to examinees on March 30, 2026. This timeline allowed NCBE to evaluate each phase of the exam lifecycle sequentially, including pre-exam preparation, exam delivery, grading, scaling and equating, and performance reporting.

The beta functioned as the bridge between development and operations.

November 3, 2025	Jurisdiction Portal opened; jurisdictions began administration setup and management, including roster upload, location assignment, and readiness monitoring
December 1, 2025	Candidate Portal opened; candidates began completing “Ready to Sit” requirements, including secure browser installation and tutorial completion
December 3, 2025	Constructed-response grader training began
January 8, 2026	Candidates approved for extended time began testing
January 9, 2026	Standard-time candidates began testing
February 4, 2026	Grader team leader and consensus calibration training began
February 6, 2026	Constructed-response grading began
February 27, 2026	Constructed-response grading ended and scaling and equating processes began
March 30, 2026	Beta performance reports delivered to examinees

The administration was intentionally designed to surface both strengths and areas for refinement. Systems were exercised under operational load, contingency protocols were tested, and administration policies were applied in real time. In limited instances, conditions were permitted that would not occur during operational administration in order to preserve data continuity and maximize learning. These design decisions were deliberate and are noted where relevant to interpretation of findings.

The sections that follow provide an overview of the beta administration, including operational processes, digital platforms, exam content, participant characteristics, and constructed-response grading. Subsequent sections outline the primary research domains evaluated and summarize descriptive findings aligned to those domains.

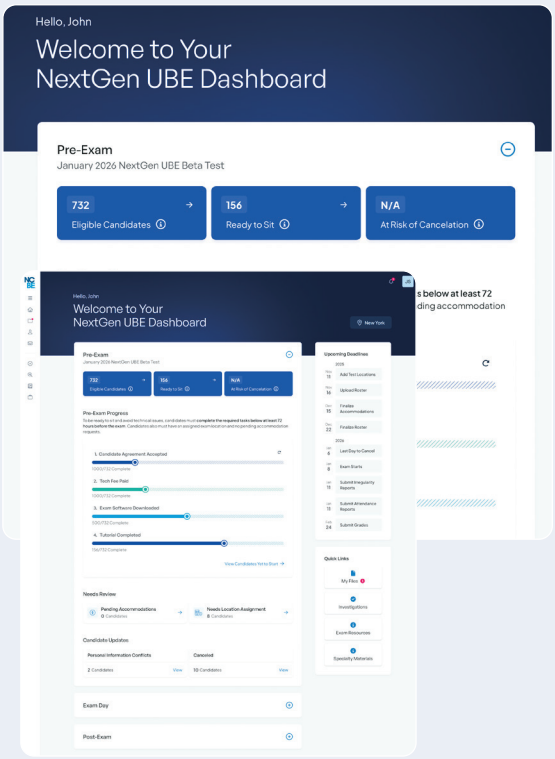
The NextGen UBE Beta Test Administration

Platforms Tested in Beta

The Jurisdiction Portal

The Jurisdiction Portal was the centralized administrative environment for jurisdictions participating in the NextGen UBE beta. The Jurisdiction Portal was designed to support the full lifecycle of exam administration—including pre-exam preparation, exam-day monitoring, and post-exam documentation—within a single integrated system.

In prior administrations jurisdictions often relied on multiple tools and manual reconciliation processes to manage exam administration. The Jurisdiction Portal consolidated candidate roster management, testing location assignment, readiness tracking, real-time monitoring, and formal reporting into a unified administrative interface. This integration reduced administrative fragmentation and improved visibility into exam readiness and delivery.



The Jurisdiction Portal provided a centralized administrative environment to support the full lifecycle of exam administration.

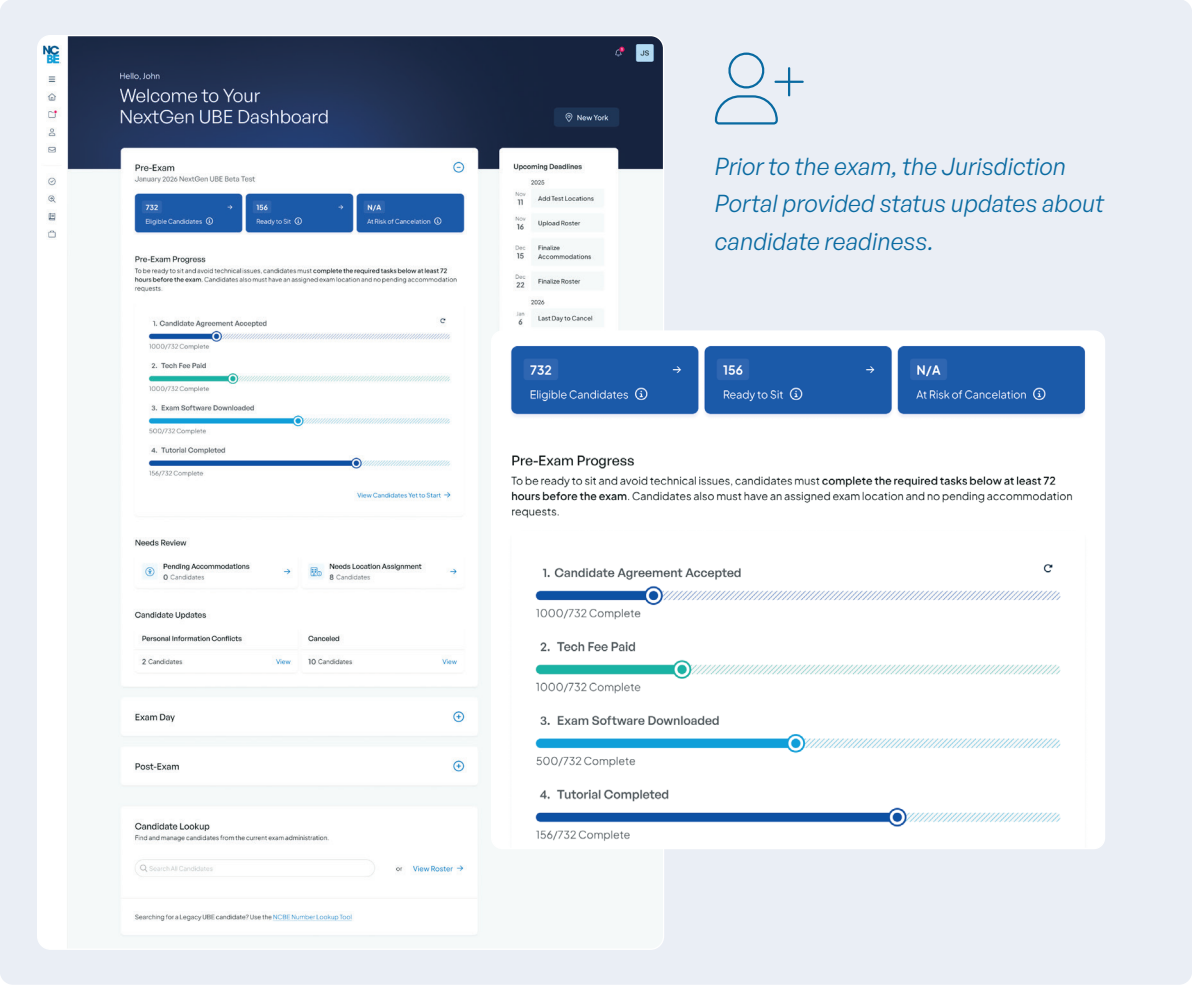
Pre-Exam Administration

During the pre-exam phase, the Jurisdiction Portal was used to:

- Upload and verify candidate rosters
- Assign candidates to testing locations
- Finalize accommodations determinations
- Monitor candidate completion of required readiness steps

Administrative dashboards provided real-time visibility into candidate readiness status, including eligibility confirmation, completion of required preparation steps, and identification of candidates at risk of cancellation. These tools allowed jurisdictions to intervene early to resolve readiness issues prior to exam day.

Centralized deadline tracking and readiness indicators reduced reliance on external tracking systems and supported coordinated preparation across jurisdictions.

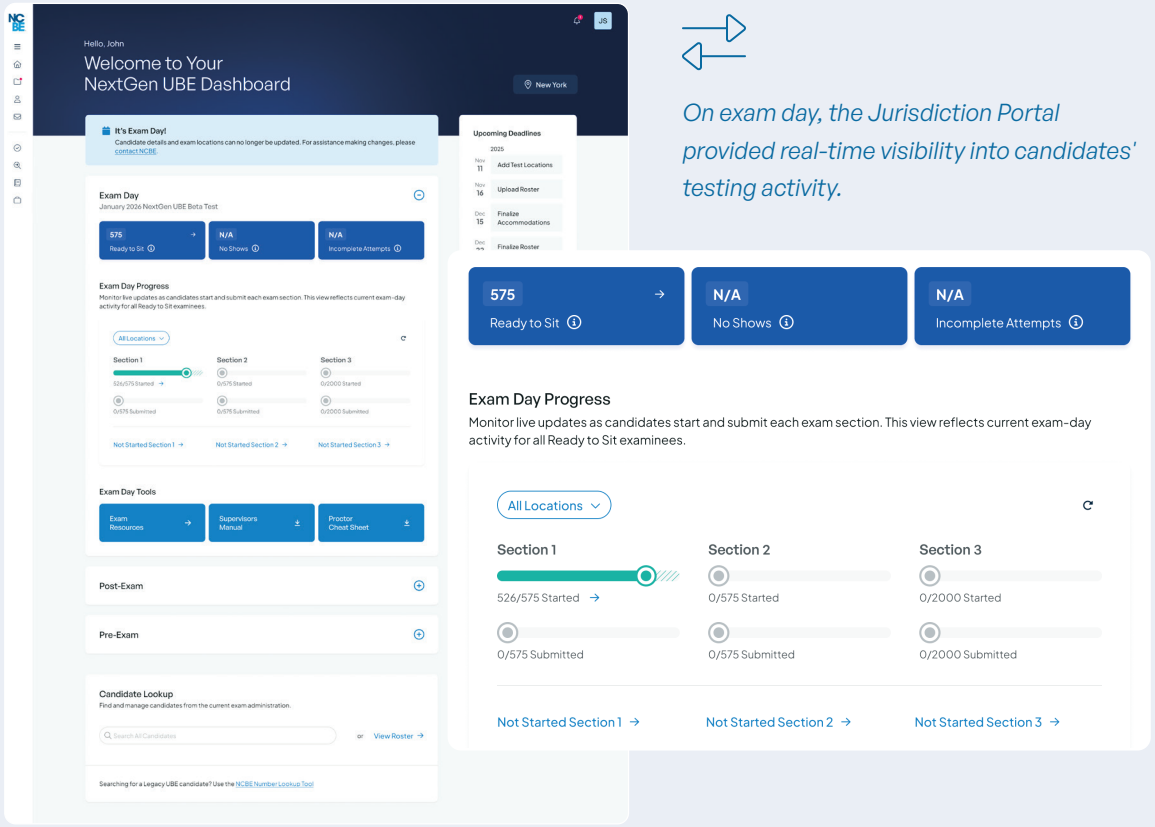


Exam-Day Monitoring

During exam delivery, the Jurisdiction Portal provided real-time administrative visibility into testing activity. Jurisdictions were able to monitor:

- Candidate login and readiness status
- Attendance and no-show designations
- Section launch and submission status
- Incomplete or interrupted testing sessions

Monitoring views could be filtered by testing location, allowing jurisdictions to oversee multiple sites simultaneously. Integrated access to administration manuals and guidance supported consistent policy implementation during live testing.

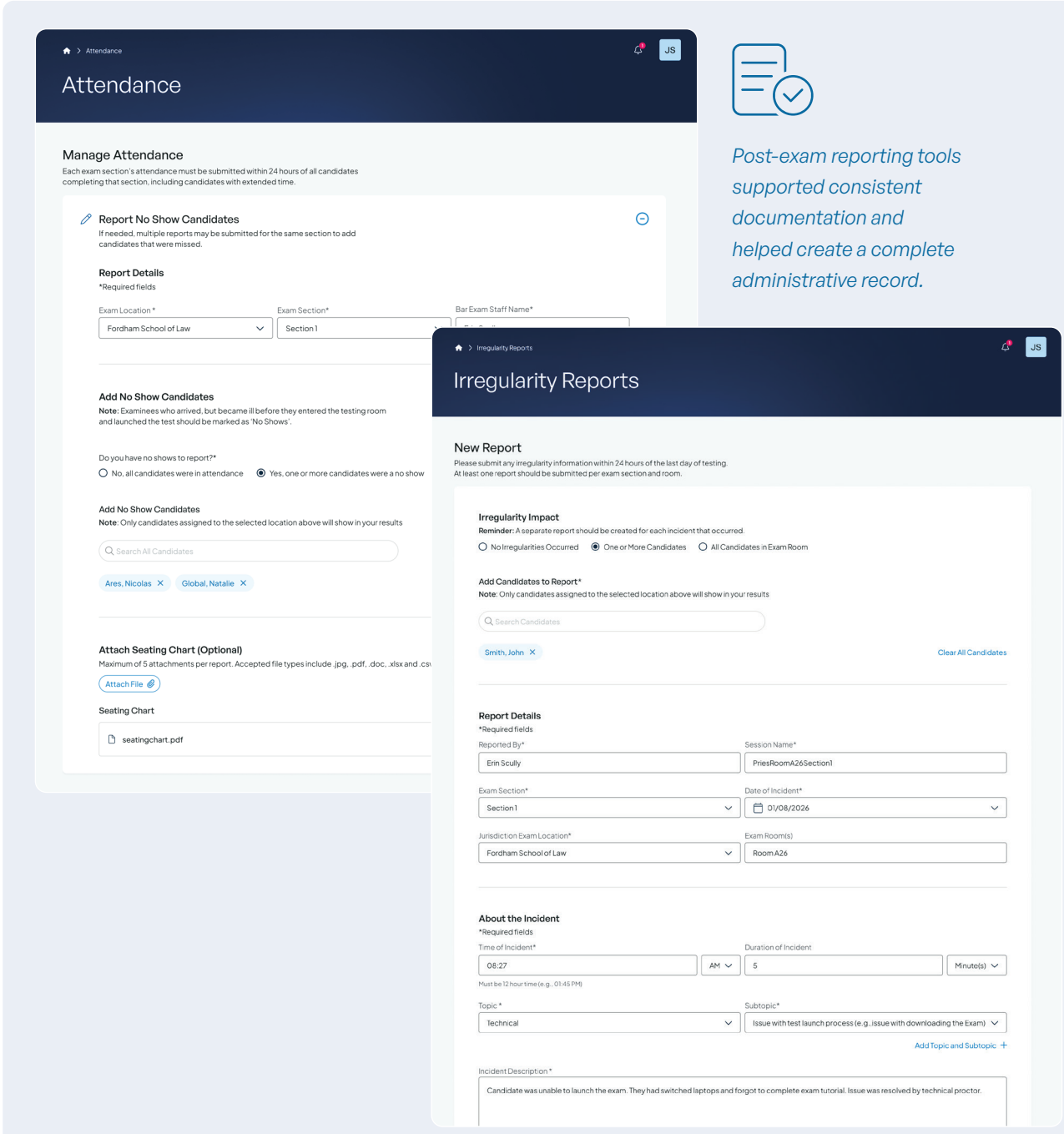


On exam day, the Jurisdiction Portal provided real-time visibility into candidates' testing activity.

Post-Exam Documentation

Following exam delivery, jurisdictions used the Jurisdiction Portal to complete attendance verification and irregularity reporting. Structured reporting forms ensured consistent documentation of technical incidents, administrative events, and environmental disruptions across jurisdictions.

Attendance records and irregularity reports were retained within the system, creating a complete administrative record supporting reconciliation, audit review, and operational analysis.



Post-exam reporting tools supported consistent documentation and helped create a complete administrative record.

Operational Impact

Across all phases of administration, the Jurisdiction Portal functioned as the primary administrative command system. It provided:

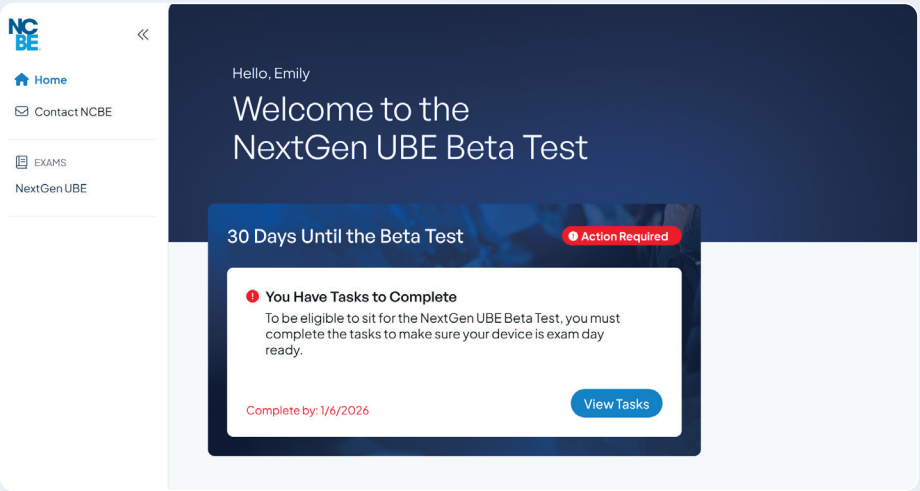
- Real-time visibility into candidate readiness and exam delivery status
- Structured workflows for attendance and irregularity documentation
- Centralized management of candidate, location, and accommodations data
- Consistent administrative oversight across jurisdictions and testing sites

The beta administration confirmed that the Jurisdiction Portal supports coordinated administration at scale while reducing reliance on manual tracking and fragmented administrative workflows.

Candidate Portal

The Candidate Portal served as the centralized candidate-facing system for the NextGen UBE beta. The Candidate Portal supported the full candidate lifecycle, including readiness preparation, accommodations tracking, and exam-day status confirmation.

In prior administrations, candidate readiness processes relied on distributed communications and manual verification. The Candidate Portal centralized readiness requirements, providing candidates with real-time visibility into preparation status and outstanding actions required prior to testing.



The Candidate Portal provided a centralized system where candidates could track their progress through the full exam lifecycle.

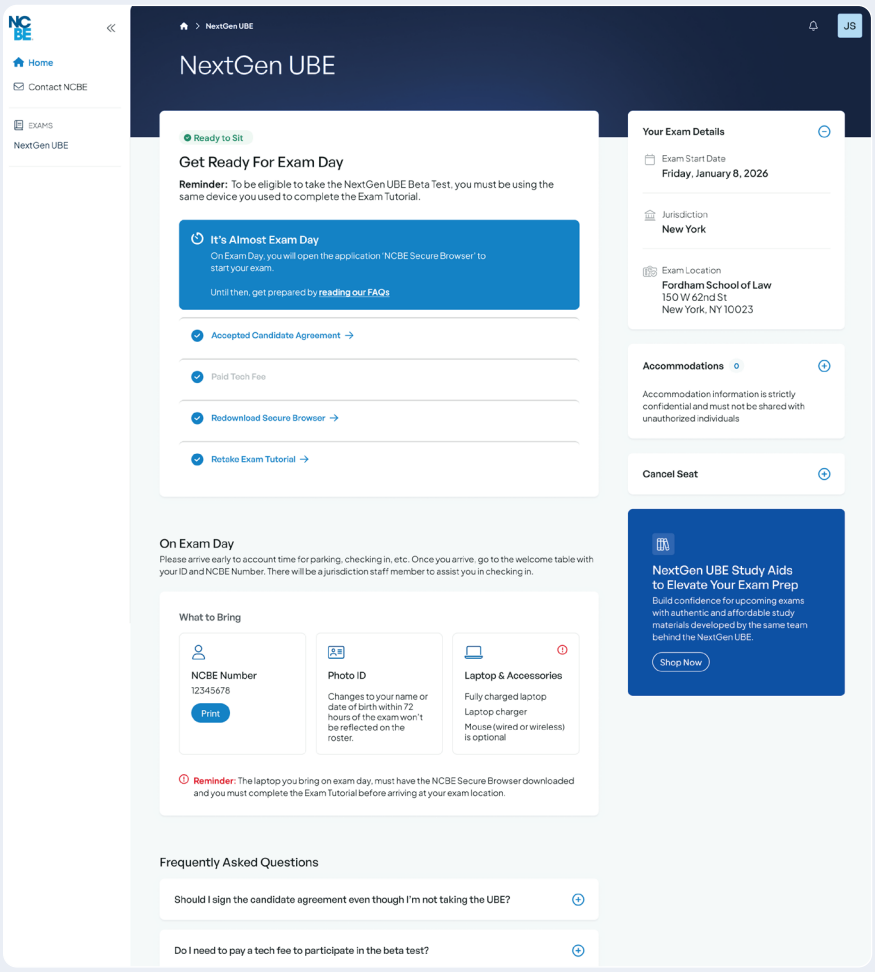
Pre-Exam Readiness

Candidates were required to complete defined readiness steps, including:

- Acceptance of the Candidate Agreement
- Payment of the technology fee
- Secure browser installation
- Completion of the exam tutorial

The Candidate Portal provided real-time confirmation of step completion and identified outstanding readiness requirements. Readiness designation required both candidate completion of preparation steps and jurisdiction confirmation of eligibility, ensuring alignment between candidate preparation and jurisdiction administration.

The portal also served as a central reference point for candidate logistics, including jurisdiction assignment, testing location, and administration date.



Candidates used the Candidate Portal to complete required pre-exam steps and ensure they were ready for exam day.

Accommodations Status and Data Integrity

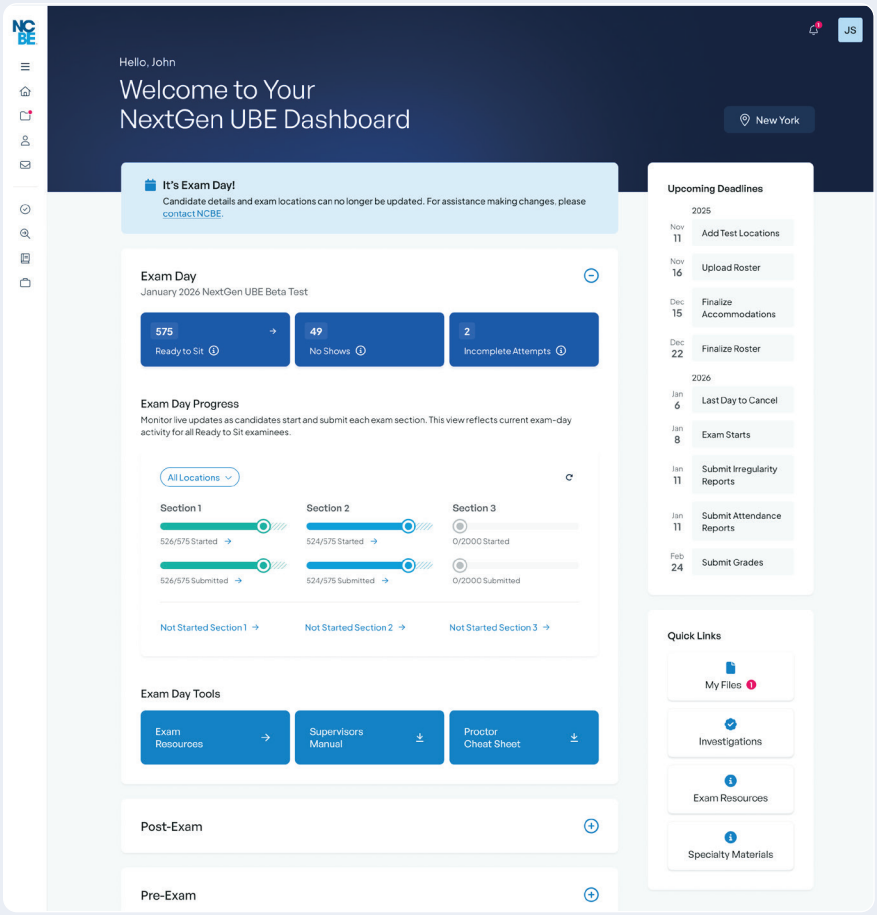
Candidates requesting accommodations were able to monitor accommodations determinations directly within the Candidate Portal. Automated validation processes identified data discrepancies, including duplicate registrations or inconsistent candidate records, enabling resolution prior to exam day.

These safeguards supported administrative accuracy and reduced the likelihood of exam-day disruptions.

Exam-Day Status Visibility

On exam day, the Candidate Portal provided confirmation of testing eligibility and response submission status without exposing exam content. Candidates were directed to launch the secure delivery platform and were able to confirm successful submission of responses following each exam section.

This submission confirmation provided candidates with assurance that their responses were successfully captured.



On exam day, candidates used the Candidate Portal to confirm testing eligibility and successful submission of their responses following each exam section.

Operational Impact

The Candidate Portal centralized candidate readiness tracking, accommodations visibility, logistics communication, and exam-day confirmation into a unified system. By consolidating these functions, the portal reduced administrative ambiguity, improved candidate awareness, and supported structured preparation at scale.

The beta administration confirmed that the Candidate Portal effectively supported candidate readiness and provided clear, real-time status visibility throughout the exam lifecycle.

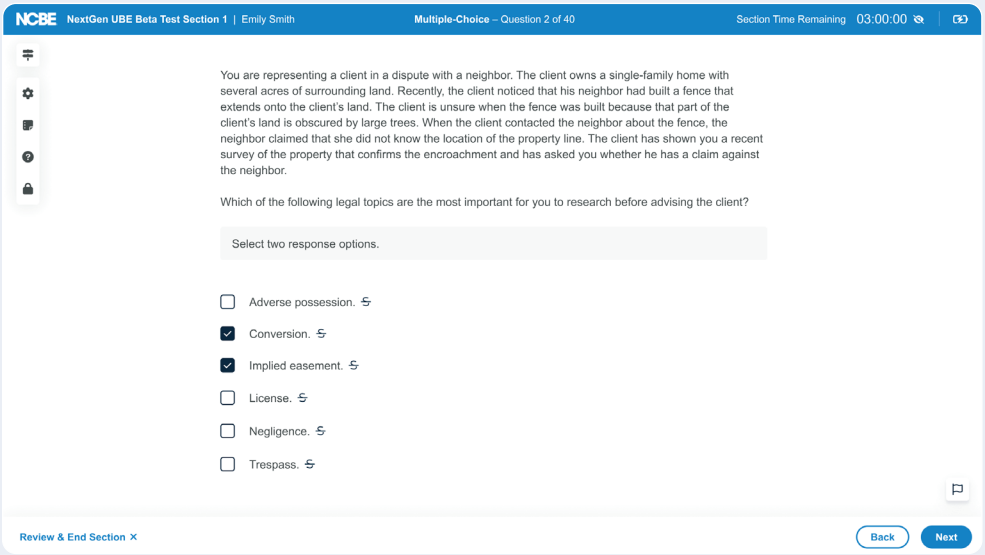
Delivery Platform

The NextGen UBE delivery platform is the secure, candidate-facing application through which all exam content was presented and all responses were captured. It served as the operational testing engine within the NextGen ecosystem and was distinct from the Candidate Portal, Jurisdiction Portal, and ITS Exam Day Portal, which supported readiness, administration, and monitoring functions respectively.

The delivery platform presented all item types within a unified interface, including:

- Standalone selected-response questions (single- and multi-select)
- Integrated question sets combining selected-response and written components
- Standard performance tasks
- Legal research performance tasks

All content was delivered within a controlled testing environment designed to protect exam integrity and ensure consistent presentation across administrations.



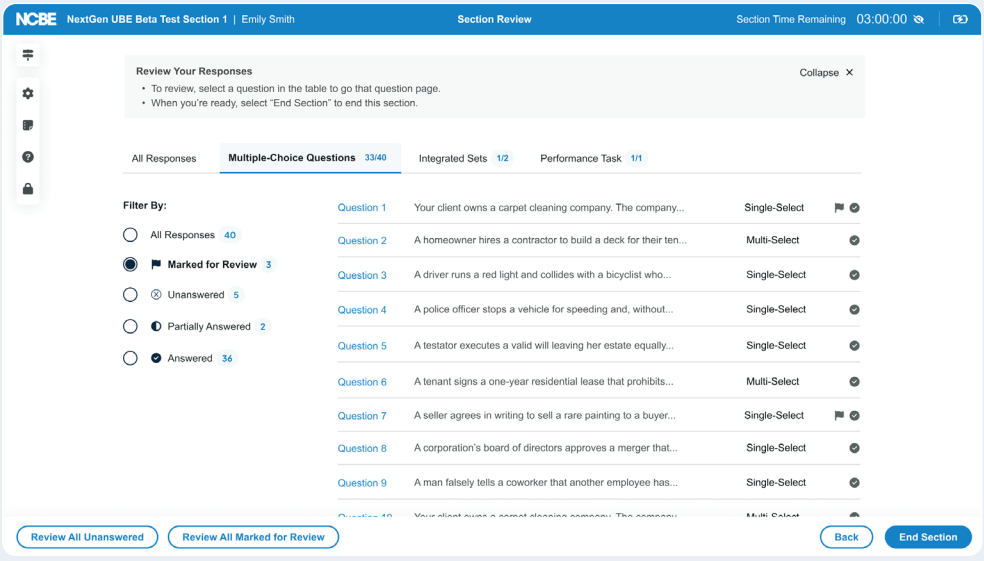
The exam delivery platform presented exam content and captured candidate responses within a single unified interface.

Content Presentation and Navigation

The platform was structured to support non-linear navigation within each section. Examinees could move between questions, access embedded case files and libraries, and return to flagged items prior to submission.

For integrated question sets and performance tasks, source materials were embedded within tabbed or split-screen layouts, allowing examinees to review case files, legal authorities, and other materials alongside response fields. Completion indicators and flagged-item markers provided visual tracking of progress throughout the section.

Structured review screens at the end of each section allowed examinees to confirm responses and review flagged items before final submission.



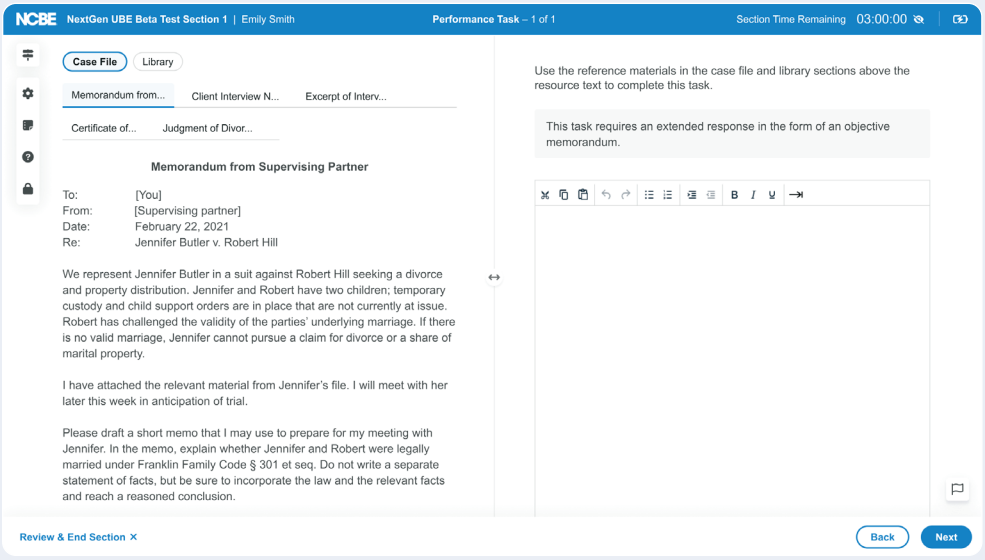
Review screens at the end of each exam section allowed candidates to confirm their answers before final submission.

Writing and Response Tools

The delivery platform included built-in tools to support written work and efficient navigation, including:

- Highlighting within source materials
- Strike-through functionality for selected-response items
- Notepad functionality
- Copy and paste capability
- Spell check
- Word count display
- Adjustable text zoom and contrast settings
- Timer visibility and toggle controls

These tools were embedded directly within the testing interface and did not require external applications.



Constructed-response questions were presented with tabbed or split-screen layouts. Built-in tools supported written work and efficient navigation.

Response Capture and Submission

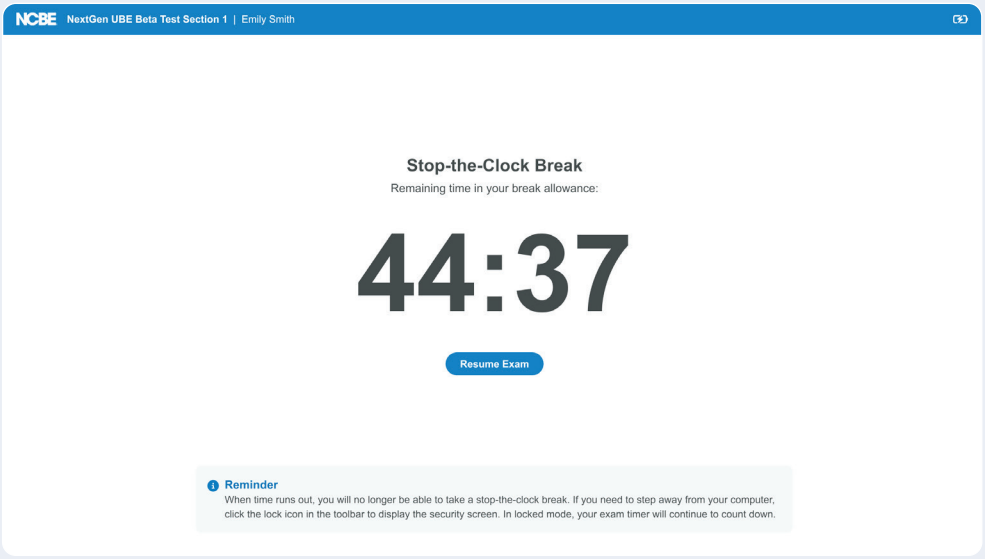
Responses were captured within the secure delivery application during active testing. At the conclusion of each section, responses were submitted through a structured submission workflow that included confirmation screens and warning modals prior to finalization.

Submission confirmation indicators provided verification that responses had been successfully uploaded and recorded.

Accessibility and Accommodations Configuration

Accessibility tools were built into the core platform and available to all examinees. These included adjustable text size, contrast controls, and consistent user interface design intended to reduce visual strain.

Accommodations-specific configurations were delivered through platform settings rather than separate systems. Approved accommodations such as extended time, stop-the-clock breaks, speech-to-text, and text-to-speech were configured within the same delivery environment used for standard administrations.



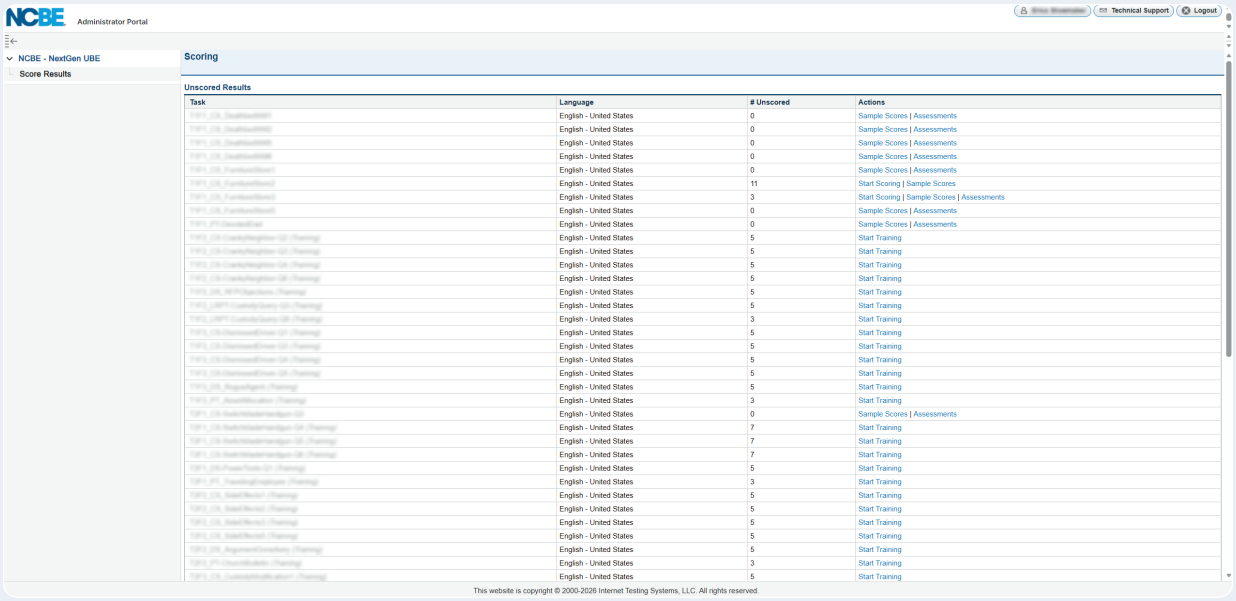
Tools for accommodated candidates, such as stop-the-clock break timers, were delivered directly through the exam platform.

Grading Platform

The NextGen UBE grading platform is a secure, digital scoring environment that was used to manage assignment, evaluation, reconciliation, and oversight of constructed-response items. It served as the operational scoring engine within the NextGen ecosystem and was distinct from the delivery platform used by examinees and from the jurisdiction and monitoring portals used for administration.

The platform supported scoring of all constructed-response components of the NextGen UBE, including integrated written responses, standard performance tasks, and legal research performance tasks.

The grading platform supported scoring of all constructed-response exam components.



Response Management and Assignment

Following exam delivery and reconciliation, constructed responses were released into the grading platform for scoring. The system assigned responses to trained graders in accordance with defined scoring models, including independent double grading and team-based oversight structures.

The platform managed:

- Secure distribution of responses to graders
- Item-specific scoring assignments
- Tracking of grading progress and completion status
- Workflow routing for responses requiring reconciliation

Assignment logic was designed to ensure that each response was evaluated independently in accordance with established scoring procedures.

Scoring Interface and Materials

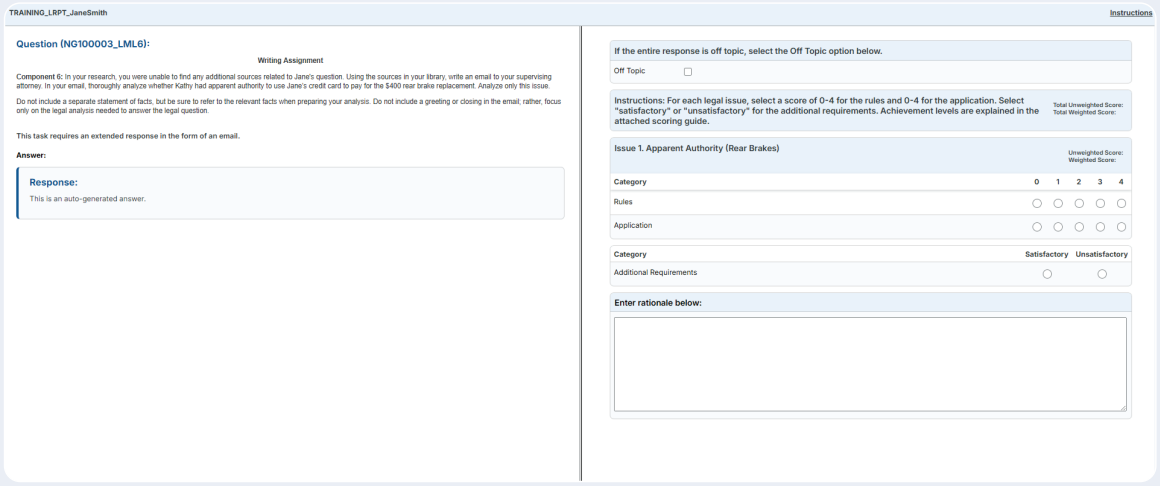
Graders accessed responses within a secure digital interface that presented examinee work alongside structured scoring materials. These materials included defined scoring criteria, performance descriptors, and representative examples developed through the content development process.

The scoring interface allowed graders to:

- Review the full examinee response
- Enter scores in accordance with defined rubrics
- Access item-specific guidance materials
- Record scoring decisions within the system

The platform maintained an audit trail of scoring activity and preserved scoring records for oversight and review.

Exam graders accessed candidate responses alongside structured grading materials in the scoring interface.



Double Grading and Reconciliation Workflows

The grading platform supported independent double grading for all constructed-response items. Each response was scored separately by two graders.

The system automatically compared assigned scores and identified differences exceeding defined thresholds. When score discrepancies met reconciliation criteria, responses were routed through structured reconciliation workflows for review and resolution in accordance with established scoring procedures.

These workflows may have involved consensus review or adjudication by designated team leaders, depending on the scoring model in use.

Oversight and Operational Monitoring

The grading platform provided administrative visibility into scoring progress and workflow status. Authorized oversight personnel were able to monitor:

- Completion rates by item and grading team
- Reconciliation volume and routing status
- Timeline adherence
- Scoring consistency indicators

This centralized monitoring structure supported operational management of large-scale grading activities while preserving the independence of individual scoring decisions.

Security and Data Integrity

All constructed responses and scoring records were maintained within a secure digital environment. Access to the grading platform was role-based and restricted to authorized graders and oversight personnel. Scoring data were stored and managed in accordance with established data governance and exam security protocols.

Content Tested in Beta

The January 2026 beta administration included three complete exam forms constructed to reflect the intended operational blueprint of the NextGen UBE. Each form followed the same structural architecture: a standalone selected-response block followed by integrated, scenario-based question sets and performance tasks incorporating both selected-response and constructed-response components. This structure reflects the NextGen UBE's emphasis on assessing applied legal reasoning and professional skills within realistic practice contexts.

Across all forms, the exam measured examinees' ability to interpret factual records, apply governing legal standards, and communicate legal reasoning through both selected-response and written formats. The beta administration provided an opportunity to evaluate the functioning of this integrated assessment model under live conditions while preserving the structural framework intended for operational launch.

Form 1 – Primary Operational Form

Form 1 served as the primary operational form for the beta administration. Its structure, content distribution, and weighting aligned with the operational blueprint scheduled for implementation.

The form opened with forty standalone selected-response questions sampling broadly across the doctrinal scope of the examination, including criminal law and procedure, civil procedure, contracts, evidence, torts, real property, constitutional law, business associations, and family law. These questions included both single-select and multi-select formats designed to assess doctrinal understanding, analytical reasoning, and the application of legal principles to factual scenarios.

Following the standalone block, the form transitioned to integrated question sets and performance tasks designed to simulate professional legal work. These included:

- A counseling question set grounded in torts and real property
- A drafting question set anchored in civil procedure and evidentiary analysis
- A performance task centered on a custody-related family law scenario

Across these question sets and tasks, examinees were required to review source materials, identify relevant legal standards, and produce structured written analysis reflecting the types of reasoning and communication expected of entry-level lawyers.

Forms 2 and 3 – Operationally Equivalent Research Forms

Forms 2 and 3 followed the same structural blueprint as Form 1, including identical sequencing, timing, weighting, and task architecture. These forms incorporated an increased concentration of family law content to support ongoing research and development.

Family law is scheduled to be added to the operational blueprint in 2028. The beta administration provided an opportunity to evaluate family law questions under live testing conditions, supporting future content integration while preserving structural consistency with the operational exam model.

Although the doctrinal anchoring of individual questions and question sets varied across forms—for example, incorporating trusts and estates, constitutional law, or criminal procedure—the overall structure and administration experience remained consistent. This approach allowed NCBE to evaluate new content areas while maintaining comparability of administration conditions.

Design Integrity Across Forms

Across all three forms, the beta administration evaluated the integrated assessment model that combines broad doctrinal sampling with practice-oriented tasks. Standalone selected-response questions provided coverage across the doctrinal scope of the examination, while question sets and performance tasks required examinees to apply legal principles, analyze factual records, and communicate legal reasoning in structured written formats.

This design reflects the NextGen UBE's focus on assessing both legal knowledge and the application of that knowledge in professional contexts.

The beta administration therefore served two complementary purposes. First, it evaluated exam forms constructed in accordance with the operational blueprint scheduled for implementation. Second, it generated research data to support ongoing content development and future blueprint evolution while preserving the structural framework intended for operational administration.

Administrative Process

The beta test was administered through a structured three-phase operational lifecycle: pre-exam preparation, exam-day execution, and post-exam reconciliation and transition to scoring. Each phase incorporated defined roles, verification checkpoints, and escalation protocols designed to support administrative consistency, data integrity, and controlled decision-making across testing sites.

This operational model reflects the administration framework planned for operational launch and was evaluated under live conditions during the beta.

Pre-Exam Preparation

The pre-exam phase focused on finalizing candidate eligibility, testing location assignments, and accommodations determinations prior to exam delivery. Jurisdictions established testing sites and managed candidate records through the Jurisdiction Portal. Jurisdictions uploaded and verified candidate rosters, assigned candidates to locations, and finalized accommodations approvals.

All candidate records were required to be finalized no later than 72 hours before exam delivery to support administrative stability and ensure readiness across testing locations.

The Exam Roster served as the authoritative administrative record, consolidating candidate identity, location assignment, accommodations status, and readiness indicators within a single system. This centralized structure reduced reliance on cross-system reconciliation and provided jurisdictions with real-time visibility into candidate readiness status.

Candidate Eligibility

Testing Location Assignments

Accommodations Determinations

Concurrently, candidates completed required readiness steps through the Candidate Portal, including acceptance of the Candidate Agreement, payment of the technology fee, secure browser installation, and completion of the exam tutorial. Readiness designation required both candidate completion of preparation steps and jurisdiction confirmation of eligibility, ensuring alignment between candidate readiness and jurisdiction administration.

Automated discrepancy-detection processes identified potential data conflicts—including mismatched personal identifiers or duplicate jurisdiction records—allowing jurisdictions to resolve inconsistencies prior to exam delivery.

By the close of the pre-exam preparation period, jurisdictions had finalized testing locations, verified candidate rosters, resolved accommodations determinations, and confirmed candidate readiness. These preparation activities supported administrative coordination and reduced the likelihood of exam-day disruption.

Exam-Day Execution

Exam-day administration procedures were designed to reflect operational launch conditions while incorporating layered monitoring, verification, and escalation protocols.

Each testing room operated under defined administrative roles, including room supervisors, proctors, and technical support personnel. This separation of responsibilities supported efficient issue resolution while maintaining exam integrity.

Prior to the launch of each exam section, physical attendance counts were reconciled with digital session status through the ITS Exam Day Portal (monitoring portal). This reconciliation confirmed alignment between physical attendance and system records before testing began.

Throughout exam delivery, jurisdictions and NCBE staff monitored candidate testing activity through real-time dashboards within the ITS Exam Day Portal. These dashboards provided visibility into session launch status, testing progress, and response submission indicators.

Escalation protocols distinguished isolated device-level issues from potential system-level concerns. Where technical disruptions occurred, structured troubleshooting procedures and backup device workflows supported continued testing. Centralized support channels allowed rapid escalation and resolution when needed.

At the conclusion of each exam section, response submission status was verified prior to candidate dismissal. Where submission discrepancies were identified, administrative staff confirmed response capture through monitoring dashboards and followed defined escalation protocols as necessary.

These verification procedures supported complete capture of candidate responses and administrative documentation of exam delivery.

Post-Exam Reconciliation and Transition to Scoring

Following exam delivery, jurisdictions completed attendance verification and irregularity reporting through structured reporting workflows within the Jurisdiction Portal. These reports documented candidate participation and any administrative or technical events observed during testing.

System-level reconciliation processes then confirmed alignment across:

- Physical attendance records
- Digital testing session records
- Response transmission and storage records

Escalated technical incidents were reviewed and resolved prior to scoring activation.

Once delivery reconciliation was complete, constructed responses were released to the grading platform, and grading assignments were initiated. This transition allowed evaluation of the full operational workflow from exam delivery through scoring activation.

Following completion of administration and scoring initiation, NCBE conducted a structured operational review examining administrative workflows, incident reports, escalation patterns, and system performance.

Findings from this review informed targeted refinements and operational readiness planning in advance of the July 2026 operational administration.

Participants and Participation

Sample Design and Participant Characteristics

Rationale for a Stratified and Representative Sample

A central objective of the beta test was to administer the exam to a sample of examinees that reflected the population of individuals who will take the exam operationally, while also supporting research and development. For a high-stakes licensing examination, sample composition is critical. A representative sample makes it possible to evaluate exam delivery, usability, and operational performance across the range of individuals expected to participate in future administrations, while also supporting subsequent analyses related to fairness, accessibility, and score interpretation.

1,500 examinees completed all sections of the beta administration

At the same time, exam development requires sufficient representation of specific candidate subgroups to support meaningful research. For this reason, the beta employed a stratified sampling design intended to approximate the characteristics of the bar-taking population while also ensuring sufficient representation of subgroups necessary for research and development. This approach supports evaluation of exam functioning, accessibility features, and operational workflows across diverse candidate populations.

The resulting sample balances two complementary objectives: reflecting the expected bar-taking population and supporting the research and development necessary prior to operational launch.

Overall Participation and Form Assignment

A total of 1,500 examinees completed all sections of the January 2026 NextGen UBE beta administration. Examinees were assigned to one of three complete exam forms.

Form assignment distribution was as follows:

- **Form 1:** 1,038 examinees (69.2%)
- **Form 2:** 232 examinees (15.5%)
- **Form 3:** 230 examinees (15.3%)

Examinees were assigned through a stratified sampling design that balanced salient candidate characteristics across forms. Ensuring that the form groups had comparable distributions of candidate characteristics enabled fair cross-form evaluation of operational delivery and workflow performance.

A subset of examinees tested under approved accommodations conditions, including extended-time administrations. If an examinee had been approved for an accommodation in the July 2025 bar exam administration, they were approved in the beta. This allowed evaluation of accommodations workflows and accessibility features within the same delivery environment used for standard administrations.

Full descriptive characteristics by form can be found in the following table.

Characteristic Distribution					
CHARACTERISTIC	**MBE**	**Beta**	**Form 1**	**Form 2**	**Form 3**
MBE SCALED SCORE					
MBE < 120.5	11.96	12.60	12.62	12.07	13.04
120.5 <= MBE < 155.5	61.85	65.33	63.87	67.67	69.57
MBE >= 155.5	26.18	22.07	23.51	20.26	17.39
GENDER					
Man	37.85	37.27	37.76	36.64	35.65
Woman	53.23	55.20	54.43	57.33	56.52
Non-binary	.77	.87	.96	.43	.87
Other	.06	–	–	–	–
No Responses	8.10	6.67	6.84	5.60	6.96
RACE/ETHNICITY					
White	48.64	41.60	43.45	40.09	34.78
Black or African American	8.92	14.47	13.58	17.67	15.22
Asian	12.63	11.40	11.08	8.62	15.65
Latina or Latino or Latine or Hispanic	8.17	11.80	11.08	13.79	13.04
Native American or Alaskan Native	.24	.13	.19	–	–
Native Hawaiian or other	–	–	–	–	–
Pacific Islander	.09	.07	–	.43	–
Multiracial	7.96	9.27	9.44	8.62	9.13
Other	1.50	1.20	1.16	.43	2.17
No Responses	11.86	10.07	10.02	10.34	10.00

Characteristic Distribution					
CHARACTERISTIC	**MBE**	**Beta**	**Form 1**	**Form 2**	**Form 3**
DISABILITY					
No	76.00	80.07	80.06	80.17	80.00
Yes	7.87	6.00	5.59	6.47	7.39
No Responses	16.13	13.93	14.35	13.36	12.61
FIRST TO LAW SCHOOL					
No	15.68	11.93	13.10	8.19	10.43
Yes	73.71	78.67	76.97	84.05	80.87
Did not go to law school	.13	.33	.48	–	–
Unsure	.54	.47	.39	.86	.43
No Responses	9.94	8.60	9.06	6.90	8.26
FIRST TO COLLEGE					
No	73.41	69.80	70.71	66.38	69.13
Yes	14.85	18.73	17.34	23.28	20.43
Unsure	.94	1.47	1.16	3.02	1.30
No Responses	10.79	10.00	10.79	7.33	9.13
ENGLISH FIRST LANGUAGE					
No	19.57	23.80	22.83	25.00	26.96
Yes	69.70	66.93	67.92	66.38	63.04
Unsure	.53	.80	.48	1.29	1.74
No Responses	10.20	8.47	8.77	7.33	8.26

Role of Form Design in Sample Distribution

The distribution of examinees across forms reflects the distinct operational and research purposes of each form.

Form 1 was constructed to align fully with the operational blueprint. Because operational forms must function consistently across the range of examinees expected to take the exam, the largest portion of the sample was assigned to Form 1. This allowed evaluation of exam delivery, administration workflows, and candidate experience under conditions closely approximating operational use.

Forms 2 and 3 followed the same structural blueprint but included additional family law content to support research and development objectives. Assignment of smaller, but sufficient, samples to these forms allowed evaluation of delivery workflows and administration processes while supporting ongoing content development.

This distribution ensured that all forms were administered under live conditions while preserving the integrity of operational blueprint evaluation.

Distribution of Examinee Performance Levels

To support evaluation across the range of examinees expected to participate in operational administrations, the beta sample included examinees representing a range of preparation and performance backgrounds.

Across forms, examinees were distributed across performance bands without concentration in a single performance range. The majority of examinees fell within the middle performance band, with meaningful representation across lower and higher performance levels.

This distribution allowed evaluation of exam delivery, platform functionality, and administration workflows across examinees with varied preparation levels.

Distribution of Candidate Characteristics

Gender

Gender representation was broadly distributed across forms. Women comprised a majority of examinees, with men also represented across all forms. Representation patterns were consistent across forms and reflected expected participation patterns for bar exam administrations.

Race and Ethnicity

The beta sample included examinees from a broad range of racial and ethnic backgrounds. Representation across forms allowed evaluation of administration workflows and delivery systems across diverse candidate populations.

These distributions support ongoing research examining exam functioning and accessibility across candidate groups.

Additional Characteristics

The beta sample also included examinees representing a range of educational backgrounds, accommodation statuses, and jurisdiction affiliations. Participation by examinees testing under accommodated conditions allowed evaluation of accommodations delivery workflows and accessibility features under live conditions.

Summary

The January 2026 NextGen UBE beta included 1,500 examinees participating across three complete exam forms administered under live, operationally representative conditions.

The stratified sample design supported evaluation of exam delivery, administration workflows, accessibility features, and candidate experience across the range of examinees expected in operational administrations. Random assignment across forms ensured that each form was administered under comparable conditions.

This sample design supported the beta administration's primary purpose: evaluating the operational delivery of the NextGen UBE ecosystem while supporting ongoing research and development in advance of operational launch.

Jurisdiction Participation

In addition to examinees, the beta included participation from jurisdiction staff who attended testing sites to observe exam administration and, in many cases, support live operational activities. Jurisdiction participation was an intentional component of the beta design, allowing jurisdictions to observe the full administration lifecycle and gain direct familiarity with NextGen administration workflows, platforms, and procedures.

64 jurisdiction representatives from 40 jurisdictions attended the beta administration

A total of 64 jurisdiction representatives from 40 jurisdictions attended the beta administration. Participants included bar admission administrators, executive directors, clerks of court, operations managers, and technology leaders responsible for bar exam administration within their jurisdictions.

Jurisdiction staff attended primarily in an observational capacity to evaluate exam-day workflows, platform functionality, and administrative coordination. Observers were able to monitor candidate check-in procedures, exam launch workflows, use of the Jurisdiction Portal and ITS Exam Day Portal, and escalation protocols supporting exam delivery.

In many cases, jurisdiction representatives also supported administration activities. Some participants served in defined operational roles, including room supervisors, proctors, and check-in personnel, while others provided support as needed during exam delivery. This participation allowed jurisdiction staff to gain direct experience with the operational model while preserving the structured administration framework.

Jurisdiction participation provided two important operational benefits. First, it allowed jurisdictions to observe and evaluate the NextGen UBE administration model under live conditions, supporting readiness for future operational administrations. Second, it allowed NCBE to receive feedback from experienced jurisdiction administrators regarding administrative workflows, platform usability, and operational procedures.

The participation of jurisdiction staff contributed to a shared operational understanding of the NextGen UBE administration model and supported coordinated preparation for operational launch.

Constructed-Response Grading

Overview of the Constructed-Response Grading Model

The beta test included large-scale implementation of the constructed-response grading model, allowing NCBE to evaluate scoring workflows, grader training and calibration procedures, and grading platform functionality under conditions approximating operational administration. The beta provided an opportunity to observe the full scoring process—from grader onboarding through score entry and reconciliation—while generating data to support subsequent psychometric analyses.

The NextGen UBE constructed-response scoring model uses a structured, criterion-based approach. Responses are evaluated against defined performance expectations and scoring criteria rather than through comparison with other examinees. This approach is intended to support consistent application of scoring standards and alignment between scoring decisions and defined performance expectations.

The beta administration included implementation of all major scoring components, including development and use of scoring materials, grader recruitment and training, calibration procedures, independent double grading, reconciliation workflows, and platform-based scoring oversight.

Scope of Beta Constructed-Response Grading

Constructed-response grading during the beta was conducted at a scale designed to approximate operational administration.

A total of 140 graders participated in the grading process and were organized into 54 grading teams. Of these:

- 38 teams operated using a consensus-based scoring model
- 16 teams operated using a team leader model with structured oversight and adjudication support

Grading covered **63** constructed-response questions, representing the full range of constructed-response question types included in the NextGen UBE. Across these questions, graders scored **64,346** individual responses.

Each response was independently scored by two graders. Where differences between scores exceeded defined thresholds, responses were routed to reconciliation workflows for review and resolution in accordance with established grading procedures.

This grading volume allowed evaluation of grading workflows, team structures, and platform-supported scoring processes under conditions similar in scale to operational administration.

Constructed-Response Questions and Scoring Materials

Constructed-response questions administered during the beta required examinees to analyze legal fact patterns, apply governing legal principles, and produce written responses reflecting professional legal work. Questions required both shorter written responses and extended written work.

Each constructed-response question was supported by structured scoring materials that defined performance expectations and guided grader judgment. These materials included:

- Defined scoring criteria
- Descriptions of expected performance characteristics
- Representative examples of responses at varying performance levels

These scoring materials were developed through established content development and review processes and were designed to support consistent application of scoring standards across graders.

Grader Recruitment, Training, and Calibration

The beta administration included full implementation of the grader recruitment and training process. Jurisdictions were invited to recommend potential graders, after which NCBE confirmed interest and completed onboarding for selected individuals.

Graders received question-specific training designed to familiarize them with the purpose of each question, the associated scoring criteria, and the application of scoring standards. Training included review of scoring materials, discussion of performance expectations, and guided practice applying scoring criteria.

Calibration activities occurred both prior to and during scoring. Initial calibration focused on establishing shared understanding of scoring criteria. Ongoing calibration activities allowed monitoring of scoring patterns and provided opportunities to clarify scoring expectations as needed.

Team leaders and consensus scoring teams provided additional support for calibration through structured review and discussion of scoring decisions.

These training and calibration processes supported consistent application of scoring procedures across grading teams.

Grading Workflow and Platform Functionality

Constructed-response grading was conducted using a dedicated digital grading platform that supported response assignment, score entry, and workflow monitoring.

The platform allowed NCBE to assign responses to graders, monitor grading progress, and manage reconciliation workflows. Platform functionality supported identification of scoring discrepancies and routing of responses requiring reconciliation.

The grading platform also provided operational visibility into grading progress and workflow completion.

The beta administration provided an opportunity to evaluate grading platform functionality under live operational conditions.

Double Grading and Reconciliation

Independent double grading was implemented for all constructed-response questions. Each response was scored separately by two graders.

Scores assigned by the two graders were then compared to identify differences exceeding defined thresholds. Responses exceeding those thresholds were routed to reconciliation workflows, where scoring differences were reviewed and resolved in accordance with established grading procedures.

This grading approach allows evaluation of scoring workflows and provides data supporting subsequent analysis of scoring consistency.

Evaluation Objectives

The constructed-response grading activities conducted during the beta were designed to evaluate the operational implementation of the grading model, including grader training, calibration procedures, grading workflows, and grading platform functionality.

The beta also generated scoring data that will support subsequent analyses of scoring reliability, consistency, and measurement characteristics.

Detailed psychometric analyses of constructed-response scoring—including reliability estimates, adjudication rates, and other statistical findings—will be presented in the forthcoming technical report.

Research Question Overview

This section outlines the primary domains of inquiry guiding the descriptive evaluation of the January 2026 NextGen UBE beta administration. These research domains were established to evaluate whether the full NextGen exam ecosystem—including technical systems, grading workflows, and administration processes—functioned as intended under live operational conditions.

The purpose of these research questions is to document operational performance, system functionality, and stakeholder experience during the beta administration. The findings presented in this report focus on system operation, administration workflows, and user experience.

Collectively, the following domains provide a structured framework for evaluating operational readiness and identifying areas for continued refinement prior to operational launch.

Technical Functionality and Capability Effectiveness

Domain Description

This domain evaluates whether the full digital exam ecosystem functioned as intended under live administration conditions. This includes the Jurisdiction Portal, Candidate Portal, ITS Exam Day Portal, secure delivery platform, and supporting infrastructure responsible for exam delivery, monitoring, response capture, and data transmission.

Evaluation Approach

Evaluation was based on system performance logs, platform monitoring data, incident and escalation records, administrative documentation, jurisdiction staff observations, and examinee feedback collected through structured surveys and post-administration review.

Research questions addressed in this domain include:

1. Did the Jurisdiction Portal function as intended to support jurisdiction administration and candidate management?
2. Did the Candidate Portal function as intended to support candidate awareness, readiness, and confidence during the beta administration?
3. Were technical performance targets met?
4. Were planning failover mitigations, redundancies, and escalation workflows effective?
5. Did accommodated administrations and embedded technical capabilities function as intended?
6. Were all examinee data captured successfully?

Constructed-Response Grading Functionality and Effectiveness

Domain Description

This domain evaluates the operational implementation of constructed-response grading, including grader recruitment and training, calibration procedures, grading workflows, reconciliation processes, and grading platform functionality.

Evaluation Approach

Evaluation was based on grading platform workflow data, grader training and calibration records, scoring assignment and reconciliation data, grading timeline monitoring, and structured feedback from graders and scoring leadership.

Research questions addressed in this domain include:

1. What was the constructed-response grading process and was it executed as intended?
2. What were the interrater reliability statistics and adjudication rates by item type on each form?
3. How did graders perceive the grading process?

Examinee Experience and Perceptions

Domain Description

This domain evaluates examinee experience with the NextGen UBE beta, including interaction with digital platforms, exam content, administration procedures, and overall usability and clarity of the testing process.

Evaluation Approach

Evaluation was based on post-administration examinee surveys, structured feedback, platform usability metrics, and observations collected during exam delivery and post-administration review.

Research questions addressed in this domain include:

1. What were examinee perceptions of the beta test overall?
2. What were examinee perceptions of each item type tested?
3. What were examinee perceptions of the administrative process?
4. What were examinee perceptions of the quality of the content tested?

Jurisdiction Staff Experience and Perceptions

Domain Description

This domain evaluates jurisdiction staff experience administering and observing the beta test, including use of administrative platforms, exam-day monitoring tools, and implementation of administration policies and procedures.

Evaluation Approach

Evaluation was based on jurisdiction staff observations, structured feedback and interviews, administrative workflow documentation, incident and escalation reports, and operational review conducted following exam delivery.

Research questions addressed in this domain include:

1. What were the overall perceptions of jurisdiction staff who attended the beta test?
2. How did jurisdiction staff report the beta administration compared to the October 2024 prototype administration?
3. What were jurisdiction staff perceptions of the ITS Exam Day Portal?
4. Did jurisdiction staff provide insight into exam-day policies being tested in the beta exam?

Research Question Findings

Technical Functionality and Capability Effectiveness

The beta administration allowed NCBE to observe technical system performance across the full exam lifecycle—from candidate readiness through exam delivery and response capture—using system performance metrics, operational monitoring, jurisdiction observation, and examinee feedback.

Across testing locations, candidate populations, and administration phases, technical systems functioned as designed and supported continuous exam delivery, response capture, and administrative oversight.

RESEARCH QUESTION 1

Did the Jurisdiction Portal function as intended to support administration and candidate management?

The Jurisdiction Portal served as the primary administrative system for managing candidate readiness, attendance verification, and administration monitoring and experienced zero downtime or technical issues.

During the beta administration, the Jurisdiction Portal was used as the primary technical solution for examinee management and exam monitoring. Due to the nature of the beta, NCBE staff functioned as "administrators" testing the system and establishing system stability. No technical issues were encountered, the system was stable throughout the process, and all validation checks against raw data provided evidence of validity.

During exam day, NCBE offered live demonstrations of the Jurisdiction Portal to all administrators who were present at the event.

Jurisdiction staff consistently described the system as clear and usable. Stephany Shipp of Illinois observed that "the testing software was very user friendly," noting that administration monitoring was "smooth and easy." Tina Bryant of South Carolina similarly reported that she "only heard good things about the testing software and the overall administration."

Mike Sullivan of Texas emphasized the administrative usability of the platform, stating that "the administrator software is easy to use."

These observations reflect the system's ability to provide centralized administrative visibility across jurisdictions and testing locations.

RESEARCH QUESTION 2

Did the Candidate Portal function as intended to support candidate awareness, readiness, and confidence during the beta administration?

The Candidate Portal served as the primary candidate readiness system during the beta administration. It functioned solidly and stably, and candidates reported it was highly intuitive and usable.

The Candidate Portal served as the primary candidate readiness system during the beta administration and functioned as designed. The platform operated stably throughout the readiness window and supported structured completion of required preparation steps. Examinees consistently described the system as intuitive and easy to use.

Survey results indicated that 98% of examinees reported understanding the steps required to prepare for and sit for the exam, reflecting a high degree of clarity regarding readiness requirements and preparation procedures within the Candidate Portal. Examinee comments reinforced these findings. One examinee noted, "I knew exactly what I needed to do before test day," while another described the setup process as "very straightforward and easy to follow."

The Candidate Portal guided examinees through readiness steps including secure browser installation, completion of the exam tutorial, and readiness verification. Real-time readiness indicators allowed candidates to confirm completion of required preparation steps prior to exam day. Jurisdiction staff observed high levels of readiness at testing sites. Ramana Pendyala from Washington noted that "applicants were all compliant in downloading the software and completing the required steps," and Jessica Crawley of Colorado observed that "examinees knew exactly where to report for check-in and what was needed."

From the examinee perspective, readiness tracking and submission confirmation features provided reassurance that preparation steps and exam responses were successfully completed. One examinee commented that the confirmation indicators made them "feel confident nothing was missing."

Collectively, these findings indicate that the Candidate Portal supported candidate awareness, reduced uncertainty prior to test day, and contributed to a structured and predictable readiness experience.

RESEARCH QUESTION 3

Were technical performance targets met?

Technical systems supported continuous exam delivery and response capture across testing locations and candidate populations. There were no system-level technical failures that interfered with exam delivery.

Technical performance targets included platform stability during live testing, successful response capture and transmission, and the effectiveness of recovery and escalation workflows under expected disruption conditions.

Of the 1,512 examinees who began testing, 1,500 (99.2%) completed all exam sections. The twelve examinees who did not complete testing either returned late to a section and were not permitted to sit (n=10), fell ill and could not return (n=1), or were asked to leave for questionable conduct (n=1). No examinee was unable to complete testing due to platform instability or delivery failure.

Although NCBE provisioned backup devices for approximately 10% of the testing population (evenly distributed across PCs and Macs), device transitions were required for a substantially smaller percentage of examinees (2.6% on Day 1 and 1.6% on Day 2). In each instance, examinees were able to resume testing and complete exam sections successfully without loss of responses.

Examinee comments aligned with these performance indicators. One examinee noted that the platform "worked smoothly throughout the entire exam," while another shared, "I didn't encounter any technical problems." A third observed that even when minor issues arose, "they were resolved quickly and didn't interrupt my testing."

Jurisdiction observers similarly described exam delivery as stable. Joey Yamanaka of Hawaii observed that "the software worked great," and Carry Pirong of Oklahoma noted that "everything worked well."

Taken together, these findings indicate that technical performance targets related to system stability, response capture, and recovery workflows were met under live administration conditions.

? RESEARCH QUESTION 4

Were planning failover mitigations, redundancies, and escalation workflows effective?

The beta administration included evaluation of system recovery and escalation procedures under both controlled and naturally occurring disruption conditions. All mitigation strategies were executed successfully and no testing was disrupted or negatively impacted.

During controlled testing scenarios—including simulated connectivity interruptions and device replacement workflows—examinees were able to resume testing without loss of responses.

During live administration, isolated connectivity interruptions occurred for a small number of examinees. In these cases, responses were preserved locally and successfully transmitted once connectivity was restored.

Backup device workflows allowed examinees to continue testing without requiring exam restart or response reconstruction.

These observations demonstrate that response preservation mechanisms and escalation workflows supported continued exam delivery and response capture during disruption scenarios.

? RESEARCH QUESTION 5

Did accommodated administrations and embedded technical capabilities function as intended?

Accommodations delivery workflows were implemented using the same delivery platform and administrative systems used for standard administrations. All accommodations capabilities functioned as intended and were very well received by examinees.

Examinees testing under extended-time and other accommodated conditions were able to complete exam sections using the secure delivery platform. Accommodation-specific timing configurations were delivered through platform configuration rather than separate delivery systems.

Jurisdiction observer Raymond Hein of Maryland noted that "the testing software worked well in the accommodations testing environment."

These findings reflect successful implementation of accommodations delivery workflows within the integrated exam delivery system.

RESEARCH QUESTION 6

Were all examinee data captured successfully?

Exam delivery systems successfully captured and transmitted examinee responses throughout the beta administration, even when offline. All data were uploaded successfully, and had this been an operational administration, all examinees would have received valid scores.

Response capture occurred continuously during exam delivery, with responses stored locally and transmitted at section completion. Submission confirmation indicators allowed candidates and administrators to verify successful transmission.

These observations reflect stable operation of response capture and transmission systems during live administration.

Constructed-Response Grading Functionality and Effectiveness

The beta administration included full implementation of the constructed-response grading model under operationally representative conditions. The beta evaluated grader recruitment and onboarding, training and calibration procedures, independent double scoring, reconciliation workflows, grading platform functionality, and operational timelines at scale. Across 63 constructed-response items and more than 64,000 individual responses, the grading process was executed using the dedicated digital grading platform in accordance with established scoring procedures.

RESEARCH QUESTION 1

What was the constructed-response grading process and was it executed as intended?

The beta demonstrated that NCBE can execute large-scale, double-blind constructed-response grading at operational scale with full workflow integrity.

The grading process was implemented as designed. A total of 140 graders were recruited through jurisdiction nominations and formally onboarded through structured training and calibration sessions prior to live scoring. All constructed responses were independently double scored. Where score differences exceeded defined thresholds, responses were automatically routed to reconciliation workflows

within the grading platform for adjudication in accordance with established procedures.

Operational monitoring indicated that all responses were successfully assigned, scored, and reconciled within the platform environment. No system-level assignment or routing failures occurred. Grading progressed within projected timelines, and workflow monitoring tools allowed leadership to track scoring velocity, reconciliation patterns, and completion rates throughout the scoring window. These findings indicate that the grading model was operationally executable at scale and functioned in accordance with design specifications.

RESEARCH QUESTION 2

What were the interrater reliability statistics and adjudication rates by item type on each form?

Preliminary analyses indicate strong interrater agreement across constructed-response item types when the process was fully implemented within a team, consistent with expectations for a high-stakes licensing exam.

Preliminary analyses of interrater agreement indicate improvement in consistency across most constructed-response item types. Early reliability estimates suggest overall interrater coefficients in the range of approximately 0.77 to 0.97, depending on item type and task complexity. Counseling and drafting tasks generally demonstrated reliability coefficients above 0.80, while extended performance tasks demonstrated coefficients within the expected range for complex written work.

The following table indicates that interrater agreement in the beta administration improved relative to the prototype. Preliminary review also suggests meaningful variation across teams, with higher agreement observed among teams that fully implemented the team leader and consensus group models. Additional analyses will further examine the extent to which consistent team alignment contributes to improved agreement rates and reduces the proportion of responses routed to reconciliation.

Final reliability and adjudication statistics will be presented in the forthcoming technical report.

Interrater Agreement Comparison

	Beta			Prototype		
ITEM NAME	Number	Overall	Percent	Number	Overall	Percent
Overall PT Stats	2,331	3,014	77%	1,956	2,069	95%
Overall LRPT Short Answer	1,314	1,501	88%	3,386	4,112	82%
Overall LRPT One-Issue PT	1,375	1,501	92%	1,830	2,075	88%
Overall CS Stats	18,048	23,155	78%	18,530	24,548	75%
Overall DS Stats	3,136	3,240	97%	3,844	4,124	93%

RESEARCH QUESTION 3

How did graders perceive the grading process?

Graders reported high levels of clarity, calibration alignment, and confidence in the scoring model.

Grader feedback collected through structured surveys and debrief discussions indicated strong acceptance of the scoring materials, calibration model, and grading platform functionality. A substantial majority reported that scoring criteria were clear and aligned to task expectations. More specifically, 88.2% of graders stated that the scoring guide materials were adequate to support scoring, and 84.3% indicated that the training set prepared them to apply scoring criteria effectively. Similarly, 82.4% rated the rubric as highly helpful, and 81.4% reported that grading notes and benchmark responses were highly helpful in supporting consistent application of scoring standards. With respect to calibration, 72.5% of graders reported that ongoing calibration activities were sufficient to help maintain alignment with scoring expectations. Collectively, these findings reflect broad confidence in the grading framework while also identifying calibration as a continued area for refinement.

Graders also reported positive experiences with the reconciliation process. A majority rated consensus-based reconciliation sessions favorably, indicating that these sessions facilitated alignment and supported a shared interpretation of scoring standards rather than creating inconsistency.

Grader Survey Results for January 2026 Beta (n=102)

Metric	Percent Positive
Scoring Guide Adequate	88.2%
Training Set Sufficient	84.3%
Rubric Highly Helpful (4–5)	82.4%
Benchmarks Highly Helpful (4–5)	81.4%
Calibration Sufficient	72.5%

Qualitative feedback primarily focused on incremental enhancements, such as requests for additional exemplar responses and expanded filtering features within the grading interface, rather than structural concerns with the scoring model.

Examinee Perceptions

RESEARCH QUESTION 1

What were examinee perceptions of the beta test overall?

Examinees reported overwhelmingly positive experiences with the NextGen UBE beta, particularly regarding usability, practicality, and alignment to professional practice.

Examinees described the NextGen UBE beta as a modern testing experience that felt usable, professionally relevant, and easier to navigate than many high-stakes exam environments. Overall sentiment was strongly positive. Approximately 76% of examinees rated their experience as very good or excellent, while only a small percentage reported a negative experience. Taken together, these ratings indicate broad acceptance of the beta experience and confidence in the digital delivery experience.

A central theme in examinee feedback was the perceived practicality of the assessment. Many examinees emphasized that the exam rewarded applied reasoning and task-based judgment rather than memorization, and that it felt closer to professional work than traditional formats. One examinee captured this shift directly: "This exam felt more practical because it tests what we will actually do." Another described it as "more logical, and not just a test of memory." Several examinees explicitly connected the exam to their future professional responsibilities, noting that it "felt more realistic with the lawyering experience."

Usability of the digital delivery platform emerged as a second dominant theme. More than 92% of examinees reported that the platform was intuitive and easy to use, and more than 94% reported that exam materials were comfortable and easy to read. Examinees frequently described the platform as stable, clear, and responsive. One examinee summarized the experience as "highly intuitive," while another shared that it was "extremely easy to follow once you began using it." Importantly, examinees repeatedly indicated that the usability of the platform reduced cognitive overhead, allowing them to focus on the substance of the exam rather than the mechanics of testing.

Examinees also compared the experience favorably to other testing platforms. Many noted that organization and interface design improved efficiency and reduced friction under timed conditions. One examinee wrote that "the organization of materials saves so much time," while another stated that the platform "reduces a significant amount of stress during testing." These comments suggest that design choices intended to reduce procedural burden were visible to examinees and experienced as meaningful during live administration.

When examinees described what worked best, they consistently pointed to specific features that supported efficient reading, writing, and navigation—including copy/paste functionality, highlighting tools, spell check, customizable display settings, and the ability to move efficiently between materials. One examinee summarized: "The platform is easy to navigate and works smoothly," while another described it as "very straightforward."

Finally, examinees frequently described the overall testing environment as calm, organized, and well-managed, noting that technology performance and operational execution contributed to a stable test-day experience. One examinee summed up the experience simply: "It worked smoothly throughout the entire exam."

Together, these findings indicate that examinees experienced the NextGen UBE beta as credible, usable, and professionally relevant, with platform usability and exam design features contributing to reduced friction during live administration.

? RESEARCH QUESTION 2

What were examinee perceptions of each item type tested?

Examinees viewed NextGen item types as practical, authentic, and reflective of real legal tasks, with strong acceptance of performance-based formats.

Across question types, examinees most often described the NextGen questions as authentic, practice-oriented, and designed to measure applied legal reasoning. Feedback consistently emphasized that the assessment required examinees to interpret information, make judgments, and communicate legal reasoning in ways that felt closer to real legal work. One examinee stated plainly that "the performance tasks felt like actual legal work," while another

explained that the exam was "much closer to what we would actually do in practice."

Performance Tasks

Performance-based tasks received particularly strong positive feedback, with examinees emphasizing the realism of the source materials and the work-product orientation of the tasks. Examinees frequently described these tasks as evaluating practical skills and professional reasoning rather than memorization. One examinee captured that distinction as: "This exam tests how you think, not just what you remember." Many examinees also noted that the digital environment supported performance task execution through document navigation, highlighting, and editing. As one examinee explained, "the organization of materials made it easier to work efficiently," reinforcing the way platform structure can directly affect the examinee experience on complex tasks.

Multiple-Choice Questions

Multiple-choice items were generally perceived as clear and aligned with applied legal reasoning. Examinees frequently noted that questions required analysis rather than rote recall and that the difficulty felt appropriate to a licensing exam context. One examinee summarized the cognitive demand succinctly: "The exam required real thinking."

Drafting and Counseling Question Sets

Drafting and counseling question sets were consistently described as relevant to entry-level practice because they required examinees to apply legal knowledge in context and communicate clearly. One examinee highlighted the authenticity of these tasks by noting they "felt like something I would actually do as a lawyer."

Overall, examinee feedback indicates broad acceptance of the question design, with particularly strong support for performance-based formats and for question types that mirror professional legal tasks.

RESEARCH QUESTION 3

What were examinee perceptions of the administrative process?

Examinees described the administrative process as organized, clear, efficient, and responsive.

Examinees generally described the administration as organized, clear, and well-managed. They reported high levels of clarity regarding pre-exam preparation and exam-day expectations, with approximately 98% indicating that they understood the Ready-to-Sit process. This suggests that readiness requirements and preparation expectations were communicated effectively and understood by examinees.

Examinees frequently described preparation and instructions as predictable and clear. One examinee noted that "the instructions were very clear," while another shared, "I knew exactly what to expect before arriving." These perceptions indicate that readiness workflows and administrative communications reduced uncertainty going into exam day.

On exam day, examinees described logistics as efficient and support as responsive when needed. When technical issues occurred, examinees generally reported that issues were resolved quickly and without meaningful disruption. As one examinee explained, "I didn't encounter any insurmountable problems navigating the test delivery platform." Many examinees also described the test-day environment as calm and organized, which is notable given the inherent stress of a high-stakes exam setting. One examinee again summarized the experience simply: "It worked smoothly throughout the entire exam."

These findings suggest that administrative processes supported a stable test-day experience and reduced procedural uncertainty during live delivery.

RESEARCH QUESTION 4

What were examinee perceptions of the quality of the content tested?

Examinees perceived the content as rigorous, relevant, and aligned with legal practice expectations.

Examinees generally perceived the content as rigorous, relevant, and aligned with expectations of professional legal practice. Approximately 70% agreed that the exam fairly assessed minimum competence, indicating broad examinee confidence in the exam's purpose and content design.

A recurring theme in feedback was that the exam felt challenging while still grounded in practical reasoning. One examinee stated that "the exam was harder but more practical," while another said, "it tests what lawyers actually do." Examinees also emphasized the cognitive demand of the tasks. One examinee captured this directly: "The questions required real thinking." Collectively, these comments reflect a perception that the exam assessed applied reasoning and professional judgment rather than relying primarily on recall.

Examinees also reported that preparation materials and tutorials accurately represented the exam experience, reinforcing perceptions of transparency and helping examinees approach the test with clearer expectations.

Overall, examinee responses indicate strong acceptance of the content and broad support for the shift toward a practice-oriented licensing assessment designed to evaluate minimum competence through applied legal reasoning.

Jurisdiction Staff Perceptions

RESEARCH QUESTION 1

What were the overall perceptions of jurisdiction staff who attended the beta test?

Jurisdiction staff who administered the beta overwhelmingly reported that the experience reflected a modern, operationally viable exam ready for real-world deployment.

Jurisdiction staff consistently described the beta administration as organized, controlled, and reflective of a modern digital exam environment operating under live conditions. Across interviews, surveys, and direct observation, jurisdiction representatives emphasized the stability of delivery systems, clarity of administration workflows, and responsiveness of operational support.

Jurisdiction representatives from 40 jurisdictions, including bar admission administrators, executive directors, clerks, and technology leaders, attended the beta administration and observed exam operations across multiple testing sites. Many noted that the beta administration reflected the pace and structure of an operational exam rather than a limited pilot or prototype environment.

Several administrators explicitly commented on the predictability of operational workflows. One jurisdiction administrator observed that

escalation pathways were clearly defined and that support staff were responsive when questions arose. Others noted that the structure of the administration—including defined roles, escalation procedures, and centralized monitoring—allowed staff to maintain oversight without requiring ad hoc intervention.

Jurisdiction staff also described confidence in the broader operational ecosystem, including candidate readiness workflows within the Candidate Portal, administration tools within the Jurisdiction Portal, and monitoring capabilities provided through the ITS Exam Day Portal. The ability to observe candidate readiness status, attendance, and submission activity in real time contributed to administrative confidence and reduced reliance on manual tracking or reconciliation.

Several jurisdiction representatives contrasted the beta experience with prior exam transitions in other programs, noting that the level of preparation, documentation, and system visibility reduced uncertainty during live administration.

Overall, jurisdiction staff described the beta administration as structured, predictable, and operationally coherent under live testing conditions.

? RESEARCH QUESTION 2

How did jurisdiction staff report the beta administration compared to the October 2024 prototype administration?

Compared to the October 2024 prototype exam, jurisdiction staff described the beta as more stable, more predictable, and operationally smoother across the full administration arc.

Jurisdiction staff consistently reported that the January 2026 beta administration reflected meaningful operational improvements relative to the October 2024 prototype, particularly in platform stability, workflow clarity, and operational coordination.

Administrators noted that delivery systems operated more predictably during the beta, with fewer unexpected disruptions and less need for real-time troubleshooting. Where the prototype administration required more active intervention to resolve technical or procedural issues, the beta administration benefited from refined workflows, improved system performance, and clearer operational guidance.

Jurisdiction representatives also reported that administration protocols were easier to follow and more fully aligned with operational conditions. Pre-exam preparation processes—including candidate roster management, readiness tracking, and accommodations workflows—were described as more streamlined and transparent. Several administrators noted that readiness indicators within the Jurisdiction Portal provided clearer visibility into candidate preparation status than had been available during prototype testing.

Improved coordination between jurisdiction staff and NCBE support teams was also noted. Administrators described escalation pathways as clearer and resolution processes as faster and more structured.

Taken together, these observations reflect iterative refinement of administration workflows and platform functionality between prototype and beta administrations, resulting in greater operational predictability and reduced administrative burden.

RESEARCH QUESTION 3

What were jurisdiction staff perceptions of the ITS Exam Day Portal?

Jurisdiction staff reported that the ITS Exam Day Portal functioned as an effective command center, increasing visibility and reducing administrative uncertainty.

Jurisdiction staff reported that the ITS Exam Day Portal provided effective real-time monitoring of exam administration, increasing visibility into candidate activity and supporting centralized oversight during testing.

The ITS Exam Day Portal allowed administrators to monitor candidate login status, exam progression, and submission activity across testing locations. This centralized visibility reduced reliance on manual attendance tracking and allowed administrators to confirm exam progression in real time.

Jurisdiction staff consistently described the monitoring interface as structured and easy to interpret. The ability to view candidate status across locations provided reassurance that exam delivery was proceeding as expected and allowed staff to identify and respond quickly to irregularities when they occurred.

Several administrators noted that centralized monitoring capabilities improved situational awareness compared to prior administration models, which often relied on manual reporting from individual testing rooms or sites.

Where feedback was provided, it focused primarily on potential workflow enhancements, such as expanded reporting options or additional filtering capabilities. These suggestions reflected opportunities for incremental refinement rather than functional limitations.

Overall, jurisdiction staff reported that the ITS Exam Day Portal supported effective exam monitoring and administrative oversight during live administration.

RESEARCH QUESTION 4

Did jurisdiction staff provide insight into exam-day policies being tested in the beta exam?

The beta provided meaningful validation of exam-day policies, with jurisdiction staff confirming both feasibility and clarity of enforcement protocols.

Jurisdiction staff reported that exam-day administration policies were implementable under live conditions and supported structured and consistent exam delivery.

Jurisdiction representatives observed implementation of identity verification procedures, candidate check-in workflows, device management protocols, and escalation procedures during live administration. These policies were described as clearly defined and operationally executable.

Administrators noted that structured escalation pathways supported timely resolution of technical and administrative issues. Where technical disruptions occurred, staff were able to follow defined procedures to resolve issues while preserving exam continuity and response integrity.

Feedback provided by jurisdiction staff focused primarily on opportunities to further clarify procedural guidance and reinforce training materials to ensure consistent application across testing sites and jurisdictions. These observations reflect normal calibration considerations in multi-jurisdiction exam administration rather than structural policy concerns.

Overall, jurisdiction staff feedback indicates that administration policies functioned as intended under live testing conditions.

Conclusion

This descriptive report documents what occurred during the January 2026 beta administration and summarizes findings related to operational delivery, system performance, and stakeholder experience. A separate technical report, scheduled for release in May 2026, will present detailed psychometric and measurement findings to support score interpretation and use, including reliability, dimensionality, subgroup analyses, constructed-response scoring statistics, and scaling and equating outcomes.

The January 2026 beta administration represents a defining transition point in the development of the NextGen UBE. For the first time, the full NextGen exam ecosystem—including the Candidate Portal, Jurisdiction Portal, ITS Exam Day Portal, secure delivery platform, grading workflows, and administration procedures—operated as an integrated system under live, operationally representative conditions.

The beta confirmed that core systems, workflows, and administration processes function cohesively across jurisdictions, testing sites, and user roles. Delivery systems supported continuous testing and response capture, candidate readiness workflows functioned as designed, and administration protocols were implemented consistently under live conditions. Jurisdiction staff were able to observe and participate in administration workflows across the full lifecycle of the exam, providing direct validation of the operational model and supporting jurisdiction readiness for operational launch.

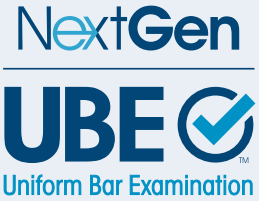
Importantly, the beta also provided the opportunity to identify targeted refinements that will further strengthen administrative efficiency, improve workflow clarity, and enhance operational consistency. These refinements reflect normal maturation of a large-scale digital assessment following live operational testing and are focused on optimizing an operationally validated system.

Planned refinements include:

- **Enhancements to reporting and export functionality within the Jurisdiction Portal,** enabling jurisdictions to access administrative data more efficiently and reduce reliance on manual reconciliation processes.
- **Expanded workflow visibility and status indicators,** allowing jurisdictions and NCBE staff to monitor candidate readiness, exam delivery, and administration events with greater clarity and confidence.
- **User interface refinements across administrative and delivery systems,** reducing administrative friction and supporting more efficient navigation during time-sensitive administration activities.
- **Additional training resources and operational guidance,** ensuring consistent implementation of administration procedures across jurisdictions and testing sites and supporting jurisdiction readiness for operational launch.
- **Continued optimization of monitoring and escalation workflows,** strengthening operational responsiveness and ensuring efficient resolution of technical or administrative issues when they arise.

Collectively, these refinements are designed to build on the strong operational foundation demonstrated during the beta administration. The focus is not redesign, but enhancing efficiency, strengthening administrative clarity, and ensuring consistent delivery as the NextGen UBE moves into operational deployment.

With the beta successfully demonstrating the operational viability of the NextGen UBE ecosystem, the work ahead centers on final optimization and preparation for the July 2026 operational launch. The NextGen UBE has now successfully transitioned from development to operational readiness, marking a major milestone in the modernization of the bar examination and the continued advancement of a licensing assessment designed to protect the public through valid, reliable, and practice-relevant measurement of minimum competence.


NextGen
UBE
Uniform Bar Examination


NCBE
National Conference
of Bar Examiners

ncbex.org