NOVEMBER 2026 / JANUARY 2027 PLANNING

IIEN Regents Standards and Stimulus Compendium

Research, form analysis, standards, exemplars, original stimuli and instructional decisions.

The North Star

A teacher should be able to begin with a standard, an unfamiliar student response, a stimulus type, a topic, or a released question and arrive at a defensible next instructional move. The route should show the evidence, the demand placed on the learner, the limits of the inference, an original task that elicits the same relationship, and a transfer task that changes the setting. Department teams should be able to compare interpretations, retain disagreements, revise a task, and check the next independent response.

The long-term system connects course standards, state design requirements, public stimuli, printed questions, official scoring evidence, original practice and anonymous local observations. Each connection carries a type and source. The current edition supplies the reference corpus, twelve course atlases, reproducible form analysis, a new bank of 72 stimulus cards, separate keys, 216 further stimulus design seeds and the earlier twelve-course exam and prep collection.

The novel cards are short reasoning tasks. They can be combined into locally reviewed sequences. The complete prior collection provides six prep sessions per course, shortened original exams and directions for assembling full-length rehearsals from public booklets. These uses require different score interpretations and exposure records.

Start from the evidence you have

With a standard, open the course atlas or filter the HTML by its code. Read the actual public mapping and the proposed alignment of the new task. With student work, begin with the scoring-evidence cases and compare the student's actual line with the requested criterion. With a topic, inspect the observed stimuli and choose a new context that preserves the evidence relationship. With a January forecast, inspect its eligible form count, comparison model and scope before deciding how much preparation time it deserves.

The HTML operates without an external service. It embeds its catalog and provides course, topic, code and source search. Its links to the original PDFs require internet access. All twelve printable atlases and student and teacher materials are in the package. CSV and JSON versions support editing, filtering and a later approved integration with the Teacher Studio.

What the corpus represents

The inherited inventory contains 139 public forms from 2022-2026. Seventy-four match the twelve scheduled January 2027 exam designs. The working item table contains 3,738 printed questions or scoring slots and 6,203 item-to-code edges. A code may represent a broad cluster, key idea or complete writing task. Multiple edges attached to one slot are different mappings of the same observation.

The course atlases expose the January, June and August denominator separately. Three Global II forms have no complete traceable official key-idea map in the inherited table. Their booklets and printed stimulus sets remain available for source-format study, while code-rate calculations use fully mapped forms. Two Biology questions receiving universal credit in June 2026 remain in the printed inventory and are excluded from content-frequency evidence.

New Algebra II, ELA and Chemistry designs have only June and August 2026 forms in this snapshot. New Physics has one June form. These courses have no observed January form in the matched design. Legacy Physics has public June forms here and a restricted January 2027 administration. Unavailable winter evidence remains missing. Prior designs offer context when their identity is explicit.

Sources are regular English releases. Large-print and translated editions can be valuable access resources, but they do not create independent forms. The package retains the inherited per-item mapping verification route and source hash register. Some official guide downloads returned errors during the refresh. Those records retain their source links and the actual verification state rather than receiving a new full-document review claim.

Evidence levels and connection meaning

An official mapping links a question or task to a state code at the published level. It supports an alignment claim of that scope. A released anchor or model response supports an interpretation of that particular scored work. An acceptable-answer criterion describes what can earn credit. A constructed rehearsal response illustrates a teaching decision. An original card's alignment is a proposed instructional connection that still needs educator review.

The Connections table uses proposed_alignment, analogous_code_example, related_scoring_evidence and cross_course_reasoning. A shared code alone does not establish that two tasks have equal difficulty or elicit the same reasoning. The detailed released-item examples include 122 inherited source-supported analyses for ELA, Algebra I and Geometry. The 36 calibration cases cover all twelve titles. Their selected examples cannot estimate how often a misconception occurs among candidates.

Every novel card names a target code, a task, an expected response, a misconception hypothesis and a transfer prompt. Science cards additionally name a practice and crosscutting concept. A short card samples part of the linked expectation. The full expectation, NYSED clarification and assessment boundary remain the authority for a course mastery decision.

Design the task around an inference

First identify the claim about student thinking. Then specify the observation that would support it and the reasoning that connects the observation to the claim. Construct a stimulus that makes that evidence necessary. Finally, define scoring criteria that distinguish correct reasoning, incomplete reasoning and an alternative explanation of an error. This local task-design process applies the assessment triangle and evidence-centered design to an instructional use. [R09, R10]

For example, the Algebra I receipt task elicits whether a student checks a pair against both constraints. A matching first bill can coexist with a failing second bill, while a proposed pair can also satisfy both. A task bank that always supplies an incorrect proposal can teach a response habit. This edition includes a new pair that satisfies both bills, so the student must actually verify it.

A science diagram should carry information needed to explain a phenomenon. A historical document should require source, context or relationship reasoning. A graph should require an inference involving its variables or scale. Define the whole demand before assigning a cognitive-complexity description. A verb, word count, number of steps or visual format alone does not establish depth of knowledge.

NYSED's science item review criteria inform review for phenomenon relevance, required reasoning, integrated dimensions, scientific accuracy and accessible wording. The cards are instructional drafts with a narrower claim than a complete state cluster. Official Regents development includes trained item writing, review of field and operational forms, scoring-material development and rangefinding. The local bank still needs educator and learner review before classroom adoption. [R02, R03]

Analyze a pattern at the correct unit

A printed slot answers how many numbered response positions carry a code. Maximum raw credits answer how much of a form's score the positions can contribute. A source set or science cluster answers how often a shared stimulus occurs. A form answers whether a feature appears at least once in an administration. The tables retain these units so a page containing six questions does not become six independent stimulus observations.

Mathematics credit shares count each slot's maximum credit once within its broad category. Category ranges are blueprint constraints, and lower or upper endpoints across categories need not sum to 100. ELA writing ratings are weighted differently from a selected response, so a count of question numbers does not describe score contribution. Task-wide writing codes are excluded from item-level frequency estimates here. Legacy science mappings overlap content, process and reference-table tags.

The observed-stimulus catalog joins history introductions and explicit science clusters to the codes on their numbered questions. That join is a source-set relationship. A cluster containing two expectations does not establish that every subpart assesses both. The science cue extractor recognizes selected words in the PDF text layer. An image-only model can be missed, and a word in a prompt can trigger a cue. Review the linked source page before writing a visual-type claim.

Probability of content appearing

For a feature observed in k of n comparable released forms, the smoothed pooled planning probability is (k+1)/(n+2), using a Beta(1,1) prior and a Bernoulli observation model. This avoids turning a small collection of all-positive or all-negative observations into certainty. If a feature appears in eight of eight forms, the model gives 9/10. It does not give the same probability to the new classroom setting used in a practice card.

The nominal 95% Wilson intervals describe uncertainty around the observed binary form proportion under an independent Bernoulli model. For a feature observed on two of two forms, that interval is approximately 34%-100%. Purposeful blueprint balancing and dependent test-development processes complicate the sampling assumption. The interval is a descriptive caution about sparse evidence, with no future-test coverage guarantee.

The topic atlas also reports a union of named codes. A slot with two named codes is counted once for that topic. A slot may still appear in more than one topic row. Those topic rows therefore cannot be added to obtain a form total. The probability targets at least one matching valid slot on a comparable form. The stimulus title, named setting, passage author, specific numbers and exact future question have no estimated appearance probability.

The January sensitivity model centers a prior on the smoothed non-January rate with strength four, then updates it with available January observations. Four is a declared model choice, not a property of NYSED development. It is retained for comparison with pooled, January-only and summer-only baselines. No model is selected because its estimate looks more appealing.

Chance and apparent seasonality

Repeated testing of many features creates opportunities for striking contrasts. The chance audit keeps each form's complete vector of feature presences intact, selects every possible subset of the same size as the observed January group, and calculates the largest absolute January-versus-other prevalence gap within each feature family. It then reports the fraction of these relabelings whose family maximum is at least the observed feature gap.

This is an exact enumeration under an exchangeable-label reference model. It preserves code co-occurrence within forms and adjusts the reference comparison for looking across the listed features in a family. It does not account for an unlimited search across all possible patterns, post-hoc feature definitions, changes in cohort, or deliberate test assembly. The reported tail fraction supports exploratory skepticism about a seasonal pattern. It does not establish a causal month effect or a confirmatory discovery.

Missing January evidence receives no chance calculation. A zero observed count remains a zero within its eligible denominator. A source-cue family, MC-code family and final-task family have separate scopes. A wide gap in a single small family is a cue to inspect the public form and seek a defensible mechanism, followed by a prospective check.

Test a forecast against a baseline

The inherited walk-forward evaluation uses eight held-out January administrations in four series. Each forecast uses strictly earlier form observations and a vocabulary of MC codes previously seen in its training period. This design cannot forecast the first occurrence of a new code. Known later corrections and present design identity make the exercise a retrospective simulation.

Each binary prediction is scored with Brier loss, (p-y)^2. Smaller is better. Count predictions use absolute error in the number of MC slots. The macro result first averages within each held-out form, then gives each form equal weight. The micro result weights every code prediction equally, giving forms with more codes more influence. Codes on the same form are dependent, so there are eight held-out form units and 120 code outcomes for each model.

The macro Brier losses are 0.15812 for January-only, 0.11353 for pooled history, 0.11332 for the shrunk January model and 0.11857 for summer-only. January-only is worse than pooled history in seven of eight forms. The tiny difference between pooled and shrunk January is insufficient evidence to declare a durable winner. This edition uses pooled history as a transparent planning baseline and displays sensitivity results alongside it.

Exact-stimulus forecasts, new-context predictions, science format cues and original-task difficulty have not passed a held-out forecast evaluation. The atlas makes these useful design hypotheses available with their scope visible. The strongest preparation choice usually combines blueprint coverage, demonstrated student need and recurring evidence demands. The planned January 2027 outcome will create the first prospective check of the frozen edition.

Guessing chance answers a different question

Under independent uniform guessing on a four-option item, the correct probability is 0.25. For 24 items, expected correct is 6 and the standard deviation is about 2.12. For an eight-item shortened form, expected correct is 2. The guessing table uses the binomial model to show tail probabilities for stated correct-count thresholds.

Real examinees use partial knowledge, eliminate choices and experience shared-stimulus dependence. Constructed responses have different scoring structures. These examples illuminate chance performance under stated assumptions. They supply no Regents pass probability, local proficiency threshold, or correction to subtract from an individual score. Matching the mix of correct option positions also supplies no evidence of test validity.

Student-level psychometrics require student-level responses

Released booklets support content study. They do not supply the examinee-by-item response matrix needed to fit item response theory, differential item functioning, test information or a local item-total relationship. The historical psychometric extract contains 20 administration summaries from the 2023 reporting cycle. Keep each reported alpha, facility, correlation and SEM attached to its exam, cohort, table and design.

Comparing a January mean facility with a June value cannot isolate form difficulty from differences in the tested groups. A reliability coefficient depends on the scores and population under the specified conditions. A historical raw-to-scale table belongs to its exact unmodified state form. The new and modified practice forms have local criterion scores and no state-score conversion.

The optional anonymous-response analyzer is executable on a future local dataset. It groups by cohort, form, rubric version and prompting condition. It reports responded, omitted and absent counts separately, facility among scored responses, complete-case corrected item-total correlation and coefficient alpha, and paired-rater exact agreement and quadratic weighted kappa. Undefined values remain empty. Complete-case selection, small samples and score distributions limit interpretation. An omission is missing evidence in this instructional dataset and is excluded from facility. An official testing rule may score it differently, so use a separately defined operational metric if needed.

No local student responses were supplied. The analyzer's verification uses labeled synthetic fixtures. It has no IRT, equating, DIF or pass-prediction implementation. Those require an appropriate design, adequate data, specialist judgment and evidence for the intended use. [R08]

Department calibration in 25 usable minutes

Spend three minutes choosing one course demand and reading the exact official criterion. Spend five minutes scoring one anonymous response independently. Spend six minutes locating the evidence line that supports each judgment and comparing it with the released sample or criterion. Spend five minutes naming a plausible misconception and an alternative explanation, such as incomplete work, unfamiliar language or misreading the prompt. Spend four minutes choosing one instructional move and one independent transfer task. Use the final two minutes to record the decision, the evidence to collect and the review date.

Use a released anchor as a comparison after the first independent judgment. Select samples spanning the distinctions the team is trying to make. Whole-score agreement can hide criterion-level disagreement, so retain both the criterion and evidence line. Where a rubric permits carried-forward error, distinguish the originating error from later consistent reasoning. The official item guide controls official credit decisions.

Local agreement summaries should include paired sample counts, the rubric version and the score distribution. A high exact agreement on uniformly zero or uniformly full-credit papers offers little evidence about borderline decisions. If a team uses weighted kappa, report its weighting rule and any undefined result caused by no score variation. Retain the original independent judgments when recording consensus.

Plan instruction for interrupted attendance

The current series uses two sessions a week for three weeks. This is a local scheduling choice designed for Burgard's short PLC periods and interrupted attendance. It incorporates retrieval, spaced revisiting, worked-example comparison and explanatory questioning informed by the IES practice guide. Outcomes for this schedule remain to be tested. [R11]

Each session should carry its own setting, quantities or source context, so a returning student can enter the current lesson. Begin with a minimal re-entry prompt, model the evidence relationship, let students compare two responses, then collect a new independent response. Revisit the same relationship in another context after a delay. Mark which work used prompts so supported performance can inform instruction without being treated as independent transfer evidence.

For English learners, preview non-target context vocabulary and use diagrams, oral rehearsal or bilingual clarification during instruction when these preserve the demand. For students with disabilities, use documented access supports and review the actual testing conditions for a rehearsal. ELA reading tasks need particular care because changing the source text can change the reading construct. A scaffold's success on one version does not establish equivalent performance on the unscaffolded version.

Generate new stimuli without teaching a template habit Preserve the target relationship while changing its surface features. For a system, vary whether a proposed pair passes both, one or neither constraint. For a graph, vary intercepts, units and a plausible domain, then require the learner to connect the representation to the question. For a source evaluation, change the research question so the same document can be useful for one purpose and limited for another.

For a science phenomenon, provide enough evidence to require a model or explanation, then change a boundary condition or introduce a competing explanation. A simulation should state its assumptions. For an argument, require evidence that both supports and constrains a claim. For historical reasoning, retain the source's identity, audience and time context when using actual sources. The new historical cards in this edition explicitly use teaching simulations. Replace them with authenticated documents when the instructional goal is primary-source sourcing or a fully developed document essay.

The 216 seeds are alternative contexts attached to the 72 finished cards. Each seed still needs a full stimulus, questions, checked key, alignment warrant, accessible representation and source review. A context that feels novel to the designer may be familiar to a learner. Novelty is therefore an exposure record and a testable transfer condition, in addition to a label on a card.

A connected resource architecture

The stable objects are ExamDesign, Standard, Form, Stimulus, Item, ScoreCriterion, ReleasedSample, OriginalTask, ResponseEvidence and TeachingDecision. Relations link a design to its blueprint, a form to its design, a stimulus to its items, an item to its published codes, a sample to its criterion, and a teaching decision to the anonymous evidence and follow-up task.

The delivered CSV and JSON registries implement the source, form, item, standard-edge, observed-stimulus, topic, original-task and teaching-connection layers. The HTML is a working reference interface over those registries. The optional local response analyzer and blank decision record support the next evidence layer. A live school data connection, automatic ingestion, shared scoring workflow and Teacher Studio deployment remain proposed extensions with separate approval and verification needs.

Source updates should be additive. Retain the downloaded bytes or their hashes, a public URL, a retrieval state, a content-design label and a locator. A changed guide creates a new source snapshot and a reconciliation record. A newly released administration becomes eligible only after confirming its design, public-release state, correction notices and mapping completeness. A failed fetch creates an access record, and leaves the last supported version identifiable.

Prospective January evaluation

Before viewing the January outcomes, preserve the current forecasts, code definitions, eligible form IDs, model parameters and hashes. After public release, verify the regular English form and corrections. Independently review the outcome code map and stimulus classifications. Score the frozen feature probabilities against the observed outcomes and retain features the old vocabulary could not predict as new outcomes.

Evaluate by course, design and feature family. Report Brier loss, count error, forecast coverage and abstentions. Compare with the frozen pooled baseline. Review whether any action based on the forecast improved coverage or student transfer, using a separately designed local study. A content forecast can be accurate while an instructional intervention is ineffective, and the reverse can also occur.

A correction to historical coding belongs in a new version with its effect on old results shown. It must not silently improve the archived forecast. When a design changes, begin a separate series and preserve older evidence for qualitative context.

Roadmap and acceptance evidence

The delivered foundation includes twelve atlases, 72 new stimuli and keys, editable topic and source registries, the retained practice series, an interactive catalog and reproducible analysis. Technical verification covers data cardinality, source joins, scoring calculations, rendered page bounds, the local HTML workflows and package integrity.

The next instructional stage is a department paper trial. Review each course's six new stimuli against the target standards, grade-level reading, timing and response space. Accept a task when independent reviewers can locate the requested evidence, agree on the key's defensible alternatives, and explain its relationship to the named code. Retain unresolved judgments in the decision record.

The next measurement stage uses an approved anonymous cohort. Collect completion and prompting status, compare independent and delayed transfer evidence, and review the adequacy of the scoring criteria. Determine sample needs from the intended precision and analysis rather than an invented universal minimum. A later cross-school or subgroup study needs a defensible sampling design and enough comparable evidence for its claims.

The next forecast stage is the frozen January 2027 comparison. The next interface stage is a concrete Teacher Studio integration review that checks search, source opening, course matching, print routes, access support and exposure tracking. Live updates, school data connections and publication require their own bounded approval.

Recovery and version history

The parent package hashes identify the Assessment Atlas, Forecast Engine, Practice Studio, Calibration pack and Exam/Prep Series. This edition adds a derivative view and new tasks. To recover from a failed local edit, retain the failed version for diagnosis and restore the last verified package. To return to the earlier workflow, retire this compendium and open the preserved parents. Connected sites and source artifacts were not modified.

The edition remains a private draft for department review. Classroom, external, public-release and active-standard decisions remain with Rachel. Runtime evidence is limited to the delivered offline analyses, response-analysis fixtures, PDF rendering and HTML workflow checks. Future integrations have no runtime acceptance claim.

Research directory

R01 is the official January 2027 schedule. R02 describes NYSED development participation. R03 supplies the science review criteria. R04-R07 locate science, history, mathematics and ELA standards. R08 supplies the professional testing standards. R09 and R10 support the assessment-design architecture. R11 informs instructional sequencing. R12-R15 locate science performance descriptions. R16-R20 locate public releases and current course resources. The Resource Register supplies exact links, use and retrieval state. Each course atlas also links its course guide and item-level source evidence.

Exact matched-design coverage

Course	Jan	June	Aug	Slots
Algebra I	2	3	3	280
Geometry	1	2	2	175
Algebra II	0	1	1	72
English Language Arts	0	1	1	52
Global History and Geography II	4	5	5	504
U.S. History and Government	3	4	4	407
Life Science: Biology	1	2	2	235
Earth and Space Sciences	1	2	2	250
Physical Science: Chemistry	0	1	1	98
Physical Science: Physics	0	1	0	50
Physical Setting/Chemistry	4	5	5	1190
Physical Setting/Physics	0	5	0	425

The source inventory also retains prior designs and retired-course forms as historical context. Slot counts include invalid printed questions, whose flags control analytical exclusion.

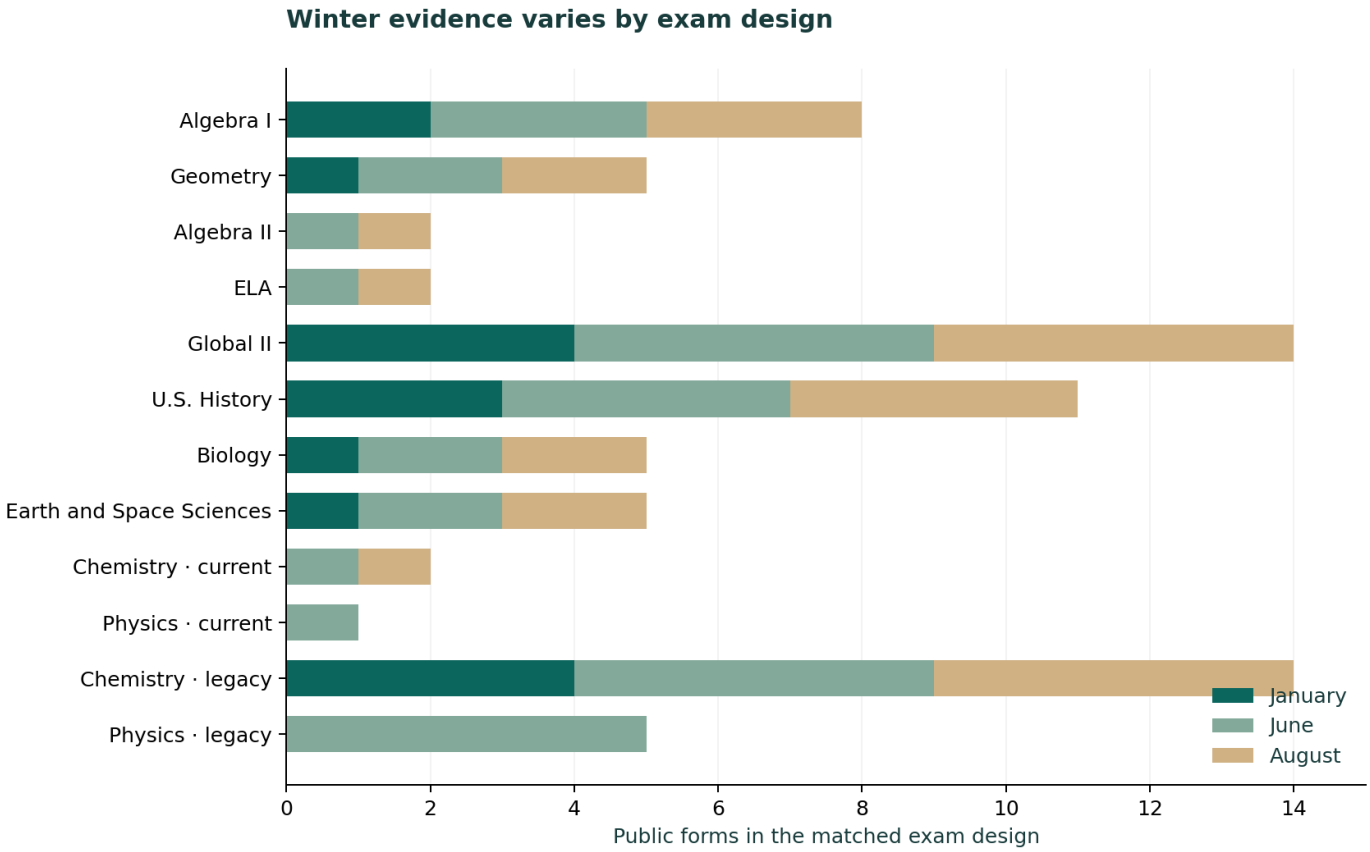
Forecast comparisons and chance illustrations

Model	Forms	Macro Brier	Count MAE
January only	8	0.1581206	1.2718858
all months	8	0.1135258	1.0778681
shrunk January	8	0.1133226	1.0779441
summer only	8	0.1185747	1.1101716

Eight held-out form units in four series. Each model has 120 established-code predictions. Lower Brier and count error are better.

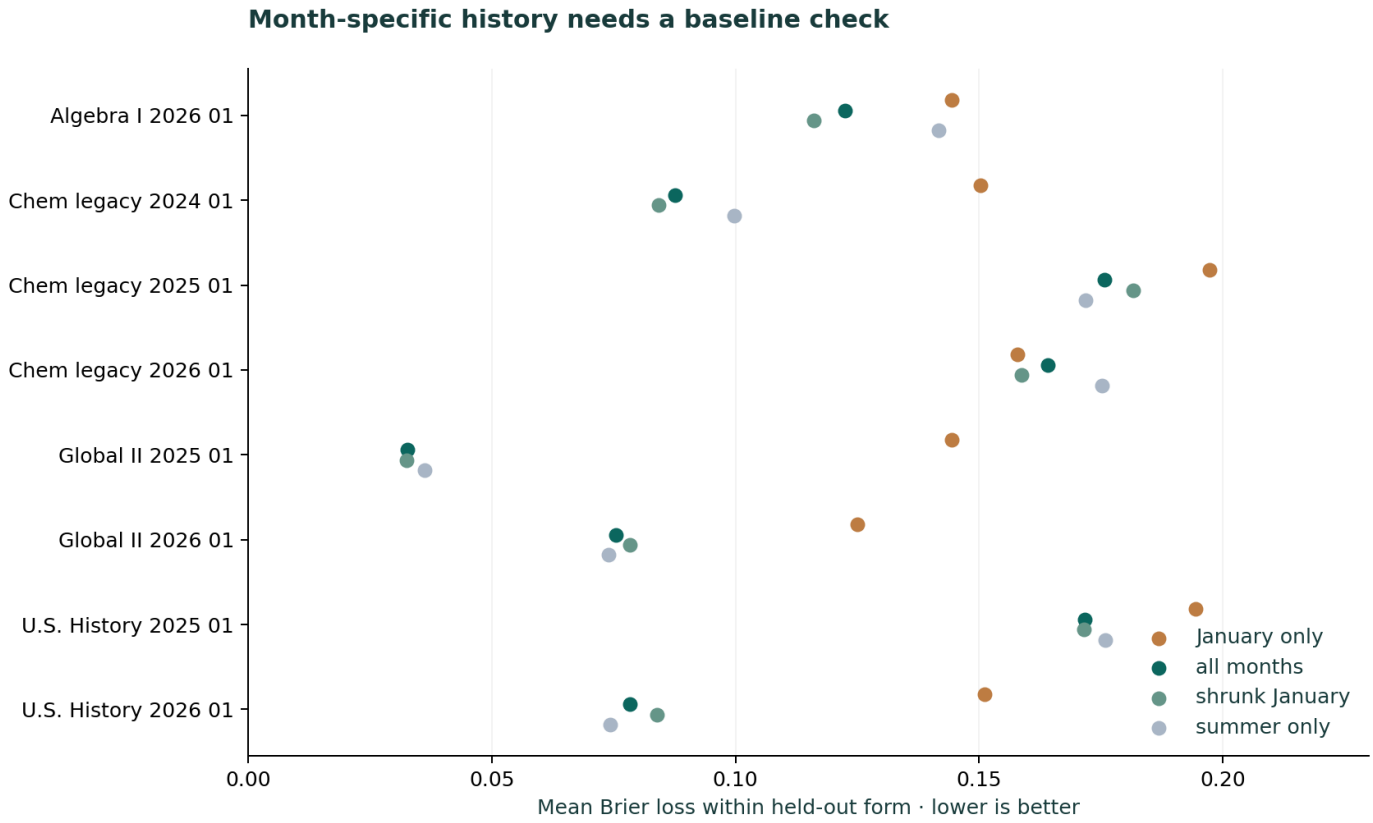
Items	Expected correct	At least	Guessing tail
8	2.0	2	0.632919312
8	2.0	4	0.113815308
24	6.0	6	0.577844798
24	6.0	12	0.00719965

How much seasonal evidence is available?



Counts are released regular English forms in the matched design, not student counts. A missing winter form remains unavailable. Legacy Physics includes June evidence only. Source data are in corpus_coverage.csv.

Compare models on the same held-out form



Each point summarizes MC-code predictions within one held-out January administration. January-only loss exceeds pooled loss on seven of eight forms. Correlated codes are not independent replications. Source data are in holdout_scores.csv.

Research sources and access state

R01 January 2027 schedule

NYSED | Schedule and restricted legacy Physics identity. Current schedule retrieved in research.

https://www.nysed.gov/sites/default/files/programs/state-assessment/regents-exams-schedule-january-2027.pdf

R02 Regents development participation

NYSED | Item writing, field and operational form review, scoring materials and rangefinding. Current indexed official page; direct open returned 502.

https://www.nysed.gov/state-assessment/participation-opportunities-regents-examinations

R03 Science Item Review Criteria

NYSED | Phenomenon relevance, integrated dimensions, scientific reasoning and review criteria. Official PDF retrieved in research.

https://www.nysed.gov/sites/default/files/programs/state-assessment/science-item-review-criteria.pdf

R04 High school science learning standards

NYSED | Performance expectations and three dimensions; consult boundaries and NY additions. Official high-school PDF indexed; full P-12 PDF also retrieved.

https://www.nysed.gov/sites/default/files/programs/curriculum-instruction/hs-science-learning-standards.pdf

R05 Social Studies Framework revised 2017

NYSED | Key ideas and historical disciplinary practices. Official PDF retrieved in research.

https://www.nysed.gov/sites/default/files/programs/curriculum-instruction/framework-9-12-with-2017-updates.pdf

R06 Mathematics implementation resources

NYSED | Design transitions and standards resources. Current indexed official page; direct open returned 502.

https://www.nysed.gov/standards-instruction/mathematics-guidance-resources

R07 ELA learning standards

NYSED | NYS 11-12 reading and writing standards; official map labels retained in the data. Official PDF indexed in research.

https://www.nysed.gov/sites/default/files/programs/standards-instruction/nys-next-generation-ela-standards.pdf

R08 Standards for Educational and Psychological Testing, 2014

AERA APA NCME | Chapters 1-4, 6, 7 and 12; intended uses, precision, fairness, development and scoring. Open-access primary PDF retrieved; references only.

https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf

R09 Knowing What Students Know, chapter 2

National Research Council | Cognition, observation and interpretation as a connected assessment design. Primary chapter retrieved; original compendium applications.

https://www.nationalacademies.org/read/10019/chapter/4

R10 A Brief Introduction to Evidence-Centered Design

Mislevy Almond Lukas, ETS 2003 | Claim, evidence, task and assembly architecture. Primary report record retrieved; report RR-03-16.

https://www.ets.org/research/policy_research_reports/publications/report/2003/hsgs.html

R11 Organizing Instruction and Study to Improve Student Learning

IES What Works Clearinghouse 2007 | Spacing, retrieval, explanatory questioning and representations. Primary practice guide indexed; local scheduling proposals are design choices.

https://ies.ed.gov/ncee/wwc/PracticeGuide/1

R12 Biology PLDs

NYSED | Interpret the performance expected at different levels; no task-to-level claim supplied. Official PDF indexed in research.

https://www.nysed.gov/sites/default/files/programs/state-assessment/life-science-biology-pld.pdf

R13 Earth and Space Sciences PLDs

NYSED | Performance expectations and resource-solution evaluation. Official PDF indexed in research.

https://www.nysed.gov/sites/default/files/programs/state-assessment/earth-space-sciences-pld.pdf

R14 Chemistry PLDs

NYSED | Science reasoning progressions. Official PDF indexed in research.

https://www.nysed.gov/sites/default/files/programs/state-assessment/physical-science-chemistry-plds.pdf

R15 Physics PLDs

NYSED | Science reasoning progressions. Official PDF indexed in research.

https://www.nysed.gov/sites/default/files/programs/state-assessment/physical-science-physics-plds.pdf

R16 Regents past examinations

NYSED | Public regular forms, keys and rating guides. Booklet sources inherited with SHA256; specific locators retained.

https://www.nysedregents.org/

R17 Algebra I current resources

NYSED | Reference sheet, PLDs and current educator guide. Official resource index retrieved in search.

https://www.nysed.gov/state-assessment/algebra-i

R18 Biology current resources

NYSED | Current guide, sample clusters and PLDs. Official resource index retrieved in search.

https://www.nysed.gov/state-assessment/life-science-biology

R19 Chemistry current resources

NYSED | Current guide, sample clusters and PLDs. Official resource index retrieved in search.

https://www.nysed.gov/state-assessment/physical-science-chemistry

R20 Physics current resources

NYSED | Current guide, sample clusters and PLDs. Official resource index retrieved in search.

https://www.nysed.gov/state-assessment/physical-science-physics