From released exams to a Burgard teaching decision

> **Priority.** Begin with one Algebra I task for January retesters. The thread has established which demands recur and how raters recognize them. The next question is which step Burgard students can complete independently.

1. Use comparable exams to select the work

The Assessment Atlas contained 139 forms and 7,241 item or scoring-slot records from 2022–26. Restricting the analysis to the designs on the 2027 schedule yielded 74 public forms and 3,738 records across 12 exam titles. Of 665 feature comparisons, 199 met the engine's threshold for a numerical planning estimate. Forecast engine and frozen evidence.

Held-out month	Earlier forms from all months won	All-month / same-month Brier ↓
January	7 of 8	.1135 / .1581
June	14 of 15	.1279 / .1624
August	15 of 15	.1056 / .1483

These 38 retrospective comparisons cover seven subjects and previously observed multiple-choice codes. Student pass probability, future essay topics and exact questions remain unknown. Published historical facility, reliability and measurement-error statistics describe their specific cohorts and forms.

2. Turn recurring demands into varied practice

All eight current Algebra I final items required a two-variable system. All five current Geometry final items required proof. All 14 Global II forms included cartoons, with 1–6 cartoon sets across the four January forms. Practice the underlying reasoning across changed contexts. The Studio supplies six sessions per title, 72 original tasks and 36 weekly released references. All-exam Practice Studio.

3. Use scoring evidence to identify the next question

The calibration guide contains 18 cases drawn from scored math, ELA and history responses and 18 based on science scoring criteria. Its 12 constructed response pairs support facilitator rehearsal. The math index locates 1,400 scored examples across 15 released sets. These selected examples establish possible reasoning needs. No Burgard student-response frequency has been established. Department calibration package.

One common task, one criterion

In January 2026 Q35, a student earned 3 of 6 credits after writing the correct system and justifying only one equation. That score opens a discussion about the evidence the question requested. It is not a scoring rule for every systems problem.

> **Criterion for the local task.** The student checks a candidate pair against the stated constraints and explains what the result establishes. One failed check disproves a solution. Confirming a solution requires every constraint to hold.

Before the PLC · Rachel and the retester teacher

Open calibration case ALG1-C1 and Studio task ALG1-1, “Two receipts.” For private rehearsal, use its supplied A/B responses. For a local finding, use authorized, deidentified attempts at the same prompt and record who had an opportunity to respond using aggregate counts. Keep independent attempts separate from prompted revisions.

During the PLC · 25 usable minutes

Time	Action and record
0–4	Read the exact prompt. Agree whether students must reject a claim, verify a solution, or show both checks. Write the criterion before viewing responses.
4–9	Independently read the released response. Identify where the student succeeded, then reveal the state commentary and compare with the higher anchor.
9–16	Two readers independently code the same anonymous work. Point to the equation, substitution, comparison or sentence supporting each judgment.
16–21	Resolve differences against the criterion. Preserve both original judgments. Keep unresolved cases and missing work visible.
21–25	Choose a single teaching move. Record the next independent task and the evidence that would change the decision.

Decision after the meeting

If the model is wrong, rehearse translating one receipt. If the model is sound but the conclusion is missing, rehearse explaining what a check proves. If students already distinguish verification from disproof, change the constraints or representation. If raters disagree about what the prompt requires, clarify the prompt and re-read before drawing an instructional conclusion.

Capture the task version, support condition, eligible attempts, exact matches and unresolved judgments. Export before closing the tab. Anonymous calibration record.

Roadmap through the next release

Stage	Owner and output	Decision gate
Independent transfer	Teacher uses new prices and a new candidate pair. An absent student can enter directly on the new task. Retain a brief aggregate readback.	Can the student choose and explain the necessary checks without the earlier cue? Record whether baseline and follow-up used comparable samples.
Department expansion	Department lead selects one case and a matched task. ELA can test evidence relevance; history can test context or sourcing; science can test claim plus item-specific evidence.	Each team agrees on what counts as evidence for its own task. A science criterion supports a hypothesis until student work is examined.
Targeted repairs	Coach and teacher use the 72-task crosswalk: 32 direct matches, 25 related, 15 needing different evidence. Repair only the selected mismatch.	Retain a task when the response can distinguish the intended skill from a competing explanation. Version changed prompts and preserve earlier observations.
January public release	Forecast owner preserves the frozen November estimates, verifies exam design and public corrections, then scores eligible outcomes in a new run.	Report hits and misses, unseen codes, missing forms and uncertainty. Restricted legacy Physics remains missing unless suitable public evidence exists.
June and August reuse	Department leads revisit local needs and the official roster. January and June list 12 titles; August lists nine. Keep legacy and current sciences separate.	Retain, revise or retire a prompt based on the matched evidence. Document human approval before wider use.

Keep two outcome records

Forecast performance concerns whether a feature occurred on the released exam. **Instructional evidence** concerns what appeared in anonymous student work on a specified task. Track agreement, independent transfer and missingness separately. Population error rates and intervention effects require stronger sampling and evaluation.

> **Current decision.** The source analysis and private rehearsal materials are available. Local calibration, classroom impact and prospective forecast accuracy are open. Use the original packages as rollback points. This roadmap proposes a sequence; it does not schedule meetings or authorize publication.