

QuestLab: Make Every Answer a Starting Point

Public whitepaper | Version 1.4 | 28 September 2026

Turning group practice into learning conversations

A useful answer does more than finish a question. It gives a learner something to explain and an educator something to build on. QuestLab puts that opportunity at the centre of a shared learning experience.

Hosts bring a topic or their own lesson materials. Learners join a room, make individual choices, examine explanations and discuss the ideas behind their answers. The host can then create a shared follow-up round around concepts the group found difficult. Practice, feedback and the next learning activity stay connected.

The practical value is immediate: a common activity for the group, feedback on accurate and mistaken selections, and material for a focused conversation. The educational ambition is to help learners explain ideas more clearly, reconsider misconceptions and use what they have learned in a new situation.

Answer. Review. Discuss. Revisit. A simple rhythm that gives the next question a purpose.

Why this approach deserves attention

QuestLab brings together design principles from learning progressions, formative assessment, retrieval practice and peer learning. Those traditions help explain why individual responses, useful explanations and purposeful follow-up belong in the same experience. This paper connects each principle to a specific product choice and sets out how the resulting learning claims can be tested.

Educators can begin with the workflow and classroom example on page 2. The research rationale and score interpretation follow on pages 3-4. A logic model and evaluation roadmap on pages 5-7 make the expected pathway and study priorities explicit.

QuestLab is built by a team with experience in teaching, STEM education research, biomedical research and learner mentoring. The product is designed around a practical educational question: once learners have answered, what should happen next?

questlab.app | hello@questlab.app

1. A useful learning cycle, in one place

Start with the lesson you want to teach

Generate a quest from a topic or use your own classroom notes and texts. Supported documents and images can be uploaded in files up to 10 MB, with text extraction and OCR for supported printed text. Hosts review the extracted text before generation. This gives the activity a starting point in the material the group is actually studying; complex diagrams and imperfect scans still need human interpretation.

Give everyone a first response

Rooms support 2-20 participants, including the host. Learners join with a nickname and room code or QR route, without creating an account. Each learner selects statements before reviewing feedback. A shared timer sets the pace. Hosts choose standard or extra time when creating a quest, and speed adds no points. Phones, tablets and computers can participate in the same session.

Use the answer to open the conversation

Each learner receives feedback on accurate ideas selected or missed and mistaken ideas selected or avoided, alongside explanations. A host can ask, "Which statement changed your mind?" or "What makes this explanation reasonable?" Learners can share that feedback and their reasoning to give the conversation a specific starting point. The host supplies the discussion and helps learners work through uncertainty.

Give the next round a reason

"Practice the gaps" generates a shared group follow-up around areas that need another look. Hosts can use it to revisit a concept after an explanation or discussion. That connection between a response and the next activity is central to the product's value.

For learners: an answer to reflect on, an explanation to question and another opportunity to practise. For hosts: lesson-based content, group progress summaries and explanations to discuss.

In a classroom: an illustrative example

A teacher brings notes on photosynthesis. Learners respond to statements about where a plant's matter and energy come from. If learners share choices that suggest confusion between sunlight as an energy source and carbon dioxide as a source of matter, the teacher can use the feedback to ask learners to explain the distinction. After discussion, a gap-focused round revisits the concept; a separate written explanation can check whether each learner can apply it to a new example.

This is a suggested use of the workflow, not a report of a completed classroom study. It shows how the game can give a teacher something specific to teach next.

2. Research that informs the design

The following connections are a focused research synthesis. They explain the design rationale; the evaluation roadmap tests how these ideas work together in QuestLab.

Make a choice before seeing the answer

Retrieval research shows that attempting to recall studied material can support later retention. Roediger and Karpicke's experiments provide a reason to make learners active participants before feedback. [9] QuestLab uses statement selection, whereas those experiments included free recall; the relevant design hypothesis is that committing to a response activates knowledge and creates a useful comparison with the explanation that follows.

Make feedback part of the learning activity

Butler and Roediger found that feedback after multiple-choice questions increased correct responses and reduced the later use of incorrect alternatives on a delayed recall test. [10] QuestLab pairs selections with explanations so learners have something to review and reconsider. In practice, a host can ask learners to identify the idea they missed and explain why an attractive alternative was wrong.

Make room for learners to explain

In undergraduate genetics courses, Smith and colleagues found that combining peer discussion with instructor explanation improved performance on subsequent concept questions more than either approach alone. [11] This supports a facilitated use of QuestLab: learners state their reasons, compare interpretations and work toward an accurate explanation with the host. The host's guidance and the quality of the conversation matter.

Let evidence shape the next activity

Black and William describe formative assessment through the use of evidence to inform decisions about learning. [5] QuestLab connects response feedback with shared gap-focused practice. The design aim is to give the host a useful basis for choosing what to revisit, and to give learners another opportunity to work with the concept.

Attend to partial understanding

Learning-progression research considers increasingly sophisticated ways of thinking within a domain. [1,4] It encourages attention to the ideas learners already use and the conceptual changes that instruction might support. QuestLab reflects that orientation through multiple statements, partial-credit feedback and further practice. A learner may recognise one important relationship while still confusing another; that pattern is worth discussing.

Together, these principles give the learning cycle a coherent purpose: surface an initial response, make the explanation useful, invite reasoning and direct the next attempt. Applying them well depends on accurate questions, a clear learning goal, suitable pacing and active facilitation.

3. Reading the learning signals

Begin with a claim, a task and an interpretation

The assessment triangle connects a model of learning, observations from tasks and an interpretation of the evidence. Evidence-centered design asks which observations would justify a particular claim about a learner. [2,3] These frameworks help QuestLab keep a practical distinction between what a response shows and what a broader judgment would require.

Recognising that one-half and two-quarters represent the same quantity is useful evidence. Explaining the relationship with a diagram and applying it to a new problem add different evidence. A host can use a quest as the beginning of that sequence and ask for an explanation or application afterward.

What the practice score tells you

0-5: A summary of the learner's accurate and inaccurate selections on the question played. Higher scores indicate stronger response performance on that task. The categories give hosts and learners a shared vocabulary for discussing a response; they are not validated stages in a domain-specific learning progression.

-1 / Not Applicable: The host has excluded a flawed or unsuitable question. It is omitted from progress calculations.

No evidence yet: A skipped or timed-out question. It is not treated as zero understanding.

Session averages: Descriptive summaries of that session. Generated quests can differ in content and difficulty, so averages should be interpreted alongside the questions and responses rather than used as a calibrated scale across different decks.

Research on partial-credit scoring supports examining how multiple-response items can preserve information that an all-or-nothing result loses. The choice of scoring rule still needs evidence for its intended use. [7] QuestLab uses a consistent rule to summarise selections against an answer key; AI helps author content and proposed keys. The product does not infer an unseen mental state directly.

Use the signal to decide what to ask next

An unexpected choice can prompt a conversation about the concept, wording or reading demand. The host can choose extra time at setup, discuss an explanation and exclude an unsuitable item. Generated content and extracted text need review because AI can produce plausible errors. [8]

QuestLab's intended use is low-stakes formative practice. Pair its feedback with learner explanations and other classroom evidence. Decisions about grades, placement or certification call for assessments validated for those purposes. [6] The score is most useful when it helps a host choose the next question to ask.

4. How the cycle is expected to work

This logic model links the experience to the changes it aims to support. Arrows represent the proposed pathway to be evaluated.

01 Resources and conditions

A defined learning goal; reviewed questions and explanations; accessible devices; sufficient time; a facilitating host.



02 Learning activities

Answer individually, review explanatory feedback, discuss reasoning and complete shared gap-focused practice.



03 Expected mechanisms

Activate knowledge; compare and reconsider ideas; articulate reasoning; focus attention and practice. Research links: [5,9-11].



04 Implementation outputs

Attempts, feedback review, discussion and follow-up completion. Collect using session records and study instruments. These describe delivery.



05 Near-term outcomes

More accurate explanations, fewer repeated misconceptions on new tasks and better-informed next teaching decisions.



06 Later outcomes

Retention at day 14 and application to unfamiliar examples. Broader conceptual development is a longer-term aim.

Delivery records show whether the cycle took place; independent tasks show whether learning changed. Discussion and explanation measures require study instruments in addition to existing session records. Findings feed back into question design, explanations, pacing and facilitation.

5. A focused evaluation roadmap

The aim is to learn which parts of the experience help, for whom and under what conditions. Five proposed studies connect the model to independent outcomes. H1 is the priority for a first study; H2-H5 investigate specific components.

H1 / Does the full cycle support durable learning?

Compare facilitated QuestLab with **structured teacher-led worksheet review**: individual concept practice followed by ordinary solution review and explanation. Match the learning goal, prior instruction, approximate task difficulty and scheduled time. Participating teachers must confirm that this represents realistic usual practice before preregistration. We predict stronger delayed performance with the full cycle. The primary outcome is an independent assessment at **day 14**, using unfamiliar applications and blind-scored short explanations, with baseline and immediate measures as supporting evidence.

H2 / Do explanations help correct mistaken ideas?

Compare explanatory feedback with correct-answer-only feedback on the same reviewed questions, with equivalent time. Measure independently coded misconception errors on unseen examples immediately and at one fixed, preregistered delayed endpoint. The prediction is fewer repeated errors with explanations.

H3 / What does discussion contribute?

Compare facilitated peer discussion with individual reflection after identical feedback, matching time and prompt demands. Collect individual explanations immediately before and after the assigned activity, then give both groups the same final reviewed explanation. Test accuracy and reasoning on unfamiliar applications. The prediction is stronger individual reasoning and application after discussion.

H4 / Does targeted group practice help where it is needed?

Compare shared gap-focused practice with broad review, matching time, question count and approximate difficulty. Measure independent improvement on concepts identified as difficult at baseline. The prediction is greater improvement on those concepts after gap-focused practice.

H5 / Can response patterns inform a better next step?

In a study configuration, randomly assign hosts to a proposed response-pattern report based on QuestLab feedback or a conventional score-and-answer summary for the same expert-reviewed learner cases, with equal planning time. Blinded raters judge proposed follow-up plans for evidence interpretation, instructional alignment, subject accuracy and a check on learning. The prediction is better alignment with independently established needs. This study concerns planned decisions; observing teaching would be a further step.

Across these studies, comparisons use independently written and reviewed tasks. Null or adverse results, or gains limited to repeated questions, would count against the relevant prediction and guide revision. A full-cycle study cannot, by itself, identify which component caused a difference.

6. Building evidence that educators can use

Current evidence and intended use

QuestLab is research-informed. As of 28 September 2026, no formal product validation or outcome study has been completed; the studies in this paper are proposed, with no active trial or preregistration reported. The literature supports the design rationale. Product-specific effects and score validity are separate questions for evaluation. The current offer is a usable workflow for shared, low-stakes practice, feedback and follow-up.

Study measures and launch requirements

The first study needs a defined setting, concept, learner group, responsible research lead and appropriate approvals and consent. Before recruitment, independently review and pilot tasks, confirm the comparison with educators, and preregister allocation, sample-size rationale, primary outcome, analysis and handling of missing data. Fix the retention point at day 14 with a prespecified scheduling window and deviation rules.

Measure the pathway as well as the endpoint. H1 should collect individual explanations before feedback and after the full cycle, using matched measurement prompts in both groups. Blind coding of accuracy, reasoning and misconceptions can track intermediate change; its association with day-14 performance would be informative without establishing causal mediation. H3's pre/post activity measure targets the discussion interval more directly.

Scoring rubrics should define concept-specific criteria, anchored examples and rules for ambiguous answers. Train independent raters on separate practice responses; blind them to condition and measurement occasion where feasible. Prespecify an appropriate agreement statistic and report uncertainty. For H5, use multiple learner cases and allow multiple defensible teaching plans. Analysis should account for classroom or room clustering, repeated decisions and missing responses. [2,3,6]

A note for education decision-makers: ESSA

ESSA's "demonstrates a rationale" category links research-supported rationale with ongoing efforts to examine effects. Federal guidance describes a well-specified logic model with a study planned as part of implementation or underway elsewhere. [12,13] This paper provides a foundation for that conversation, not an ESSA rating or endorsement. A particular funding decision requires the applicable evidence and documented study arrangements; section 1003 school-improvement funding requires one of the higher three evidence categories. [12]

A practical next step

Begin with a concept learners have already encountered. Run a facilitated quest, ask learners to explain one choice and use their responses to select the next activity. For an implementation or evaluation conversation, contact hello@questlab.app.

References: assessment and design

Selected research informing the product rationale and evaluation approach.

1. National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8*. R. A. Duschl, H. A. Schweingruber, & A. W. Shouse (Eds.). National Academies Press. Chapter 8, Learning Progressions. DOI: 10.17226/11625. [Official chapter](#).
2. National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. J. W. Pellegrino, N. Chudowsky, & R. Glaser (Eds.). National Academies Press. Chapter 2. DOI: 10.17226/10019. [Official chapter](#).
3. Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* (ETS RR-03-16). DOI: 10.1002/j.2333-8504.2003.tb01908.x. [ETS record](#). An accessible 2004 version is [CSE Report 632](#).
4. Alonzo, A. C., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education*, 93(3), 389-421. DOI: 10.1002/sce.20303. [Publisher abstract](#).
5. Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31. DOI: 10.1007/s11092-008-9068-5. [Institutional record](#).
6. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association. Chapters 1-3. [Official full text](#).
7. Betts, J., Muntean, W., Kim, D., & Kao, S.-C. (2022). Evaluating different scoring methods for multiple response items providing partial credit. *Educational and Psychological Measurement*, 82(1), 151-176. Published online in 2021. DOI: 10.1177/0013164421994636. [Full article](#).
8. National Institute of Standards and Technology. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile* (NIST AI 600-1). DOI: 10.6028/NIST.AI.600-1. [Official report](#).

References: learning and evaluation

9. Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249-255. DOI: 10.1111/j.1467-9280.2006.01693.x. [Publisher record](#).
10. Butler, A. C., & Roediger, H. L., III. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, 36(3), 604-616. DOI: 10.3758/MC.36.3.604. [Article record](#).
11. Smith, M. K., Wood, W. B., Krauter, K., & Knight, J. K. (2011). Combining peer discussion with instructor explanation increases student learning from in-class concept questions. *CBE-Life Sciences Education*, 10(1), 55-63. DOI: 10.1187/cbe.10-08-0101. [Full article](#).
12. United States Code. 20 U.S.C. section 7801(21), definition of evidence-based, including (A)(iii) and (B). [Official statute](#).
13. U.S. Department of Education. (2016). *Non-regulatory guidance: Using evidence to strengthen education investments*. See pp. 8 and 12 on rationale and study efforts. [Official guidance](#).

About this paper

Prepared by the QuestLab research and education team. Version 1.4, 28 September 2026. This is a focused research synthesis and proposed evaluation roadmap, rather than a systematic review. Product descriptions reflect functionality at publication; component-study configurations are proposals.

Public research and proprietary design

Public documentation explains intended use, score meaning and evidence. Future study reports should disclose methods, findings and limitations, including null results. Internal authoring instructions, production code and content-selection methods remain proprietary.

Continue the conversation

Explore the experience at questlab.app. Read the public research page and download the one-page brief at questlab.app/research. For educator, school or research enquiries: hello@questlab.app.