



Dialogic Assessment for Strengthening Critical Reasoning in Large Classes

Dr. Emily Carter

Associate Professor, University of Melbourne, Australia

Abstract

Large university classes create a persistent educational tension between instructional scale and the development of higher-order reasoning. Conventional quizzes and examinations can efficiently assess factual understanding across hundreds of students, yet they frequently provide limited opportunities for learners to articulate assumptions, defend claims with evidence, engage with counterarguments, revise reasoning, and respond to alternative interpretations. Dialogic assessment offers a potentially scalable alternative by embedding assessment within structured cycles of questioning, peer explanation, challenge, feedback, and reasoned revision. Evidence from higher education indicates that well-designed classroom dialogue can elicit critical-thinking processes, while regulated dialogic peer assessment can improve feedback quality and critical thinking. Active-learning research further demonstrates that student engagement through discussion and problem solving can improve achievement across large undergraduate courses.

This paper develops a structured Dialogic Critical-Reasoning Assessment Model designed for large-enrolment higher-education classrooms. The framework organizes assessment around four reasoning dimensions: evidence evaluation, claim justification, counterargument handling, and reasoned revision. It combines individual commitment, peer dialogue, structured reasoning prompts, whole-class synthesis, rapid formative feedback, and concise reflective reassessment. Because no classroom dataset was supplied, the study adopts a conceptual-methodological design and reports an explicitly simulated comparison among conventional individual quizzes, question-plus-think-pair-share assessment, peer-reasoning assessment, and structured dialogic assessment.

The simulated composite critical-reasoning scores increase from 41 under conventional individual quizzing to 93 under the fully structured dialogic model. These values are illustrative and must not be interpreted as empirical learning gains. The study concludes that dialogue becomes educationally valuable when students are required not merely to speak but to make reasoning visible, compare competing claims, justify judgments, respond to criticism, and revise positions. In large classes, digital polling, peer assessment, carefully designed prompts, and sampled instructor synthesis can distribute the work of formative assessment without eliminating academic oversight.

Keywords: dialogic assessment, critical reasoning, critical thinking, large classes, peer assessment, formative assessment, peer instruction, higher education, argumentation, assessment for learning



1. Introduction

Higher education is expected to cultivate students who can move beyond remembering disciplinary information toward evaluating evidence, recognizing assumptions, comparing competing explanations, constructing defensible judgments, and revising conclusions when stronger reasons become available. These intellectual practices are central to what is commonly described as critical thinking or critical reasoning. Yet the design of many large-enrolment courses creates a structural difficulty: instructional systems optimized for efficient information delivery and rapid grading may provide relatively little opportunity for students to make their reasoning visible. Freeman and colleagues' meta-analysis of 225 studies in undergraduate STEM education found that active-learning approaches improved examination performance and reduced failure rates relative to traditional lecturing, illustrating the broader value of moving students from passive reception toward active intellectual engagement.

Dialogue is particularly relevant because reasoning is frequently strengthened when learners must explain why they hold a position rather than merely indicate which answer they prefer. Cui and Teo's higher-education study examined a 15-week dialogic instructional program and identified discursive moves such as opening up, branching out, and tossing questions back to students as productive in drawing out critical thinking. Nussbaum's integration of argumentation theory and dialogic pedagogy similarly emphasizes that dialogue helps expose implicit premises and enables learners to examine arguments dialectically rather than treating reasoning as an isolated internal activity. Dialogue therefore has particular potential for assessment because it permits instructors and peers to observe not just conclusions but the justificatory processes through which conclusions are reached.

Assessment changes the educational significance of dialogue. An ordinary classroom discussion may be engaging while leaving reasoning quality largely unexamined. Dialogic assessment, as conceptualized in the present study, refers to a structured assessment process in which students are required to articulate a claim, provide reasons or evidence, encounter alternative perspectives, evaluate counterarguments, and revise or defend their position. Cheng and colleagues' study of regulation scripts for dialogic peer assessment found that structured prompts can support meaningful dialogue while improving feedback quality, critical thinking, and trust. More recent work on peer feedback likewise indicates that scripts, prompts, criteria, and explicit feedback structures can improve the educational quality of peer-assessment processes.

The challenge is scalability. Individual oral questioning and personalized instructor feedback are feasible in a seminar of twenty students but become difficult when an instructor is responsible for a lecture hall containing hundreds of learners. Peer instruction offers one response by distributing explanation among students. Research-based reviews describe Peer Instruction as an evidence-supported strategy in which students answer conceptual questions, discuss their reasoning with peers, and reconsider their answers. A 2024 study of large first-year mathematics courses likewise reported improvements after peer discussion, including stronger immediate and later performance on related questions compared with a non-discussion condition. Online peer assessment provides another scalable mechanism: randomized experimental work in an introductory statistics setting found that students benefited academically from engaging in peer assessment, demonstrating that assessment labor can be distributed without reducing learning to automated scoring alone.



The present study builds on these traditions but shifts attention from discussion as pedagogy toward dialogue as assessment evidence. Its central argument is that large classes do not necessarily require a choice between scalable assessment and rich reasoning. Carefully structured dialogic cycles can use individual responses, peer comparison, reasoning prompts, digital polling, representative sampling of student arguments, and reflective revision to generate evidence of critical reasoning at scale. The model does not assume that all student talk is productive, nor does it claim that peer dialogue should replace expert assessment. Instead, it identifies specific design conditions under which dialogue can reveal and strengthen evidence evaluation, justification, counterargument, and revision while preserving instructor responsibility for disciplinary accuracy and assessment validity.

2. Objectives of the Study

2.1 To Develop a Scalable Dialogic Assessment Architecture

The first objective is to design an assessment sequence that can preserve dialogic reasoning even when individualized instructor questioning is impractical. The proposed sequence integrates an initial individual judgment, peer comparison, required explanation, counterargument exchange, class-level synthesis, and reasoned revision. This structure draws conceptually on Peer Instruction, dialogic peer assessment, and dialogic feedback research, all of which treat interaction as a mechanism through which students clarify and reconsider understanding.

2.2 To Operationalize Critical Reasoning as Observable Assessment Evidence

The second objective is to distinguish critical reasoning from general participation. Students are evaluated conceptually across four observable dimensions: evidence evaluation, claim justification, counterargument handling, and reasoned revision.

The approach assumes that high-quality reasoning is visible when learners can identify relevant evidence, establish a defensible connection between evidence and conclusion, respond substantively to alternative interpretations, and modify their position when criticism reveals a weakness. Research connecting dialogic classroom moves with critical thinking and studies of dialogic peer assessment support the value of making these reasoning practices explicit rather than assuming that discussion automatically produces them.

2.3 To Examine the Feasibility of Dialogic Assessment in Large Classes

The third objective is to develop a resource-sensitive implementation model suitable for courses in which faculty cannot provide extensive individual oral feedback.

Large-class active-learning and peer-instruction evidence suggests that peer discussion, conceptual polling, and distributed feedback can be implemented at scale, although quality depends substantially on question design and instructional implementation. The present framework therefore uses technology for response collection and sampling rather than treating technology as the source of reasoning itself.



3. Review of Literature

3.1 Dialogic Teaching and the Development of Critical Reasoning

Dialogic pedagogy differs from simply allowing students to talk. Its educational value lies in the use of discourse to expose interpretations, connect ideas, test reasoning, invite alternatives, and develop understanding collaboratively. Cui and Teo's study of university dialogue showed that particular teacher moves encouraged students to elaborate thinking and extend reasoning rather than merely provide short answers. Heron's work on developing a dialogic stance in higher education similarly links reasoning and justification moves with the development of critical thinking through classroom discourse.

Argumentation provides a useful theoretical foundation for dialogic assessment because critical reasoning becomes more demanding when learners must defend a claim against plausible alternatives. Nussbaum argues that everyday arguments frequently contain implicit premises and that dialogue can expose assumptions requiring examination. This means that assessment questions designed only to identify the correct answer can underestimate reasoning quality. Two students may select the same response while reaching it through very different evidentiary and inferential processes.

Recent higher-education evidence continues to support the potential relationship between dialogic pedagogy and critical-thinking development. A 2026 quasi-experimental study at Gambella University examined the impact of dialogic pedagogy on students' critical-thinking disposition and reported positive effects associated with the intervention. Such evidence should be interpreted cautiously because critical-thinking disposition, demonstrated reasoning performance, and disciplinary learning are distinct constructs. Nevertheless, the growing literature supports the proposition that structured dialogue can create conditions in which learners practice explanation, criticism, and reconsideration rather than treating knowledge as fixed information.

A central limitation remains that not all dialogue is intellectually productive. Rapanta's investigation of classroom dialogue quality emphasizes that dialogicity depends on the structure and quality of participation rather than the simple amount of classroom talk. For assessment purposes, this implies that dialogue must be organized around explicit reasoning moves. Students need to know whether they are being asked to provide evidence, identify a hidden assumption, distinguish correlation from causation, generate a counterexample, rebut an objection, or revise a conclusion. Without such structure, participation can become socially active but cognitively shallow.

3.2 Dialogic Formative Assessment and Peer Feedback

Formative assessment becomes dialogic when evidence of student understanding is not treated as a terminal judgment but as the beginning of an exchange through which understanding can be reconsidered. Steen-Utheim and Wittek's work on dialogic feedback conceptualizes feedback as an interactive learning process rather than one-way transmission from instructor to student. A 2025 scoping review by Nedrehagen and colleagues similarly identifies dialogic formative feedback as a promising approach for deeper learning, self-regulation, and stronger student–teacher relationships, while also demonstrating continuing variation in how dialogic feedback is conceptualized and implemented.

Peer assessment is particularly useful in large classes because students can participate in both producing and evaluating reasoning. Cheng and colleagues' dialogic peer-assessment study showed that regulation scripts can direct learners toward more meaningful feedback exchanges and can positively



affect critical thinking and classroom trust. Shao and colleagues subsequently reported that regulation scripts in peer feedback could promote critical thinking and cognitively substantive feedback comments. The implication is important: peer interaction scales assessment only when peers are given intellectual tools for evaluating one another's reasoning.

Research on online peer assessment provides additional evidence that judging other students' work can itself become a learning activity. Sun and colleagues' randomized controlled study in introductory statistics found significant achievement benefits from peer assessment and emphasized its potential for providing scalable personalized feedback in large classes. Zhan and colleagues' meta-analytic work on online peer assessment similarly examined its relationship with higher-order thinking, highlighting peer assessment as a mechanism through which learners analyze criteria, compare solutions, and formulate feedback.

The literature therefore suggests that the value of peer assessment lies not merely in reducing instructor marking workload. Students who evaluate another argument must interpret criteria, detect weaknesses, compare reasoning structures, and justify a recommendation. In a dialogic model, the original author can then challenge the feedback or revise the argument. The resulting cycle—judgment, explanation, challenge, and revision—creates richer evidence of reasoning than a single-response assessment.

3.3 Scaling Reasoning Through Peer Instruction and Active Learning

Large-class instruction has long used conceptual questioning and peer discussion to create interaction without requiring the instructor to converse individually with every learner. Vickrey and colleagues' review of research-based Peer Instruction describes a sequence in which students encounter conceptual questions, commit to responses, discuss reasoning, and reconsider their answers. Tullis and Goldstone's work on Peer Instruction further shows that peer discussion combined with instructor explanation can improve learning from conceptual questions.

Recent evidence from large first-year mathematics courses provides direct relevance to the present topic. Vozzo, Johnson, and Tuke compared quiz-style questions with and without peer discussion and reported stronger immediate, one-week, and semester-end performance following peer instruction. These findings suggest that scalable dialogue can contribute to learning even where enrolment makes seminar-style interaction impossible.

The larger active-learning literature reinforces this conclusion while also cautioning against treating any one technique as universally effective. Freeman and colleagues found robust advantages for active learning across undergraduate STEM contexts, including large classes, although effects varied across settings and implementations. Theobald and colleagues subsequently found that active learning can narrow achievement gaps for underrepresented students when implemented at sufficiently high intensity, emphasizing the importance of inclusive design rather than merely adding isolated activities.

The research gap addressed by the present paper lies at the intersection of these traditions. Peer Instruction demonstrates scalable dialogue, peer assessment demonstrates scalable evaluation, and dialogic pedagogy demonstrates the relationship between discourse and reasoning. What remains less developed is a single assessment architecture that deliberately combines these mechanisms to produce observable evidence of critical reasoning in large classes. The proposed framework addresses this gap.



4. Research Methodology

4.1 Research Design

The study adopts a conceptual-methodological design with simulation-based comparative analysis. No actual students were recruited, no university examination results were analyzed, and no intervention was conducted in a live classroom. The numerical results reported in Section 5 are synthetic values generated solely to illustrate how alternative assessment designs could be compared.

The methodological framework was constructed through synthesis of research on dialogic teaching, critical thinking, peer assessment, formative feedback, Peer Instruction, active learning, and large-class participation. The design is particularly informed by evidence that structured classroom talk can elicit critical thinking, that dialogic peer-assessment scripts can improve reasoning-related outcomes, and that peer discussion can be scaled through conceptual-question formats.

4.2 Dialogic Critical-Reasoning Assessment Model

The proposed assessment cycle contains six stages:

- Individual commitment requires every learner to formulate an initial answer before hearing peers, preserving independent reasoning and reducing immediate conformity.
- Reason statement requires a short justification, evidence reference, or inferential explanation.
- Peer comparison pairs or groups students with different or potentially competing interpretations.
- Counterargument exchange requires learners to identify at least one weakness, alternative explanation, assumption, or objection.
- Reasoned revision allows students to maintain or change their original position but requires them to explain why.
- Instructor synthesis uses aggregated responses and sampled arguments to address misconceptions, compare reasoning quality, and clarify disciplinary standards.

Table 1: Proposed Dialogic Assessment Framework for Large Classes

Assessment Stage	Student Reasoning Requirement	Scalable Implementation Mechanism	Evidence Generated
Individual Commitment	Select or formulate an initial position independently	Digital poll, short-response platform, paper response card	Initial judgment
Evidence/Reason Statement	State why the position is defensible	One-sentence justification or evidence tag	Claim–evidence connection
Peer Dialogue	Explain reasoning and compare competing answers	Pair or triad discussion	Oral reasoning and alternative perspectives



Assessment Stage	Student Reasoning Requirement	Scalable Implementation Mechanism	Evidence Generated
Counterargument	Identify a weakness, objection, assumption, or counterexample	Structured dialogue prompt	Critical challenge
Reasoned Revision	Retain or change position with explanation	Second poll plus revision statement	Metacognitive and evidential revision
Instructor Synthesis	Compare representative reasoning patterns	Sampled anonymous responses and whole-class debrief	Disciplinary clarification and feedback

4.3 Simulated Critical-Reasoning Assessment

Four assessment configurations were simulated:

- Conventional Individual Quiz – students independently select or provide answers with minimal dialogic revision.
- Question + Think–Pair–Share – students answer individually, discuss briefly with a partner, and reconsider the answer.
- Peer-Reasoning Assessment – students additionally evaluate the reasoning of peers against explicit criteria.
- Structured Dialogic Assessment – students complete the full claim–evidence–counterargument–revision cycle.

The simulated scores are not standardized critical-thinking-test results and should not be interpreted as expected percentage gains in actual classrooms.

5. Analysis

5.1 Comparative Reasoning Performance Across Assessment Designs

Table 2: Simulated Critical-Reasoning Scores Across Large-Class Assessment Models

Assessment Design	Evidence Evaluation	Claim Justification	Counterargument Handling	Reasoned Revision	Composite CRDI
Conventional Individual Quiz	14	12	8	7	41
Question + Think–Pair–	18	17	14	13	62

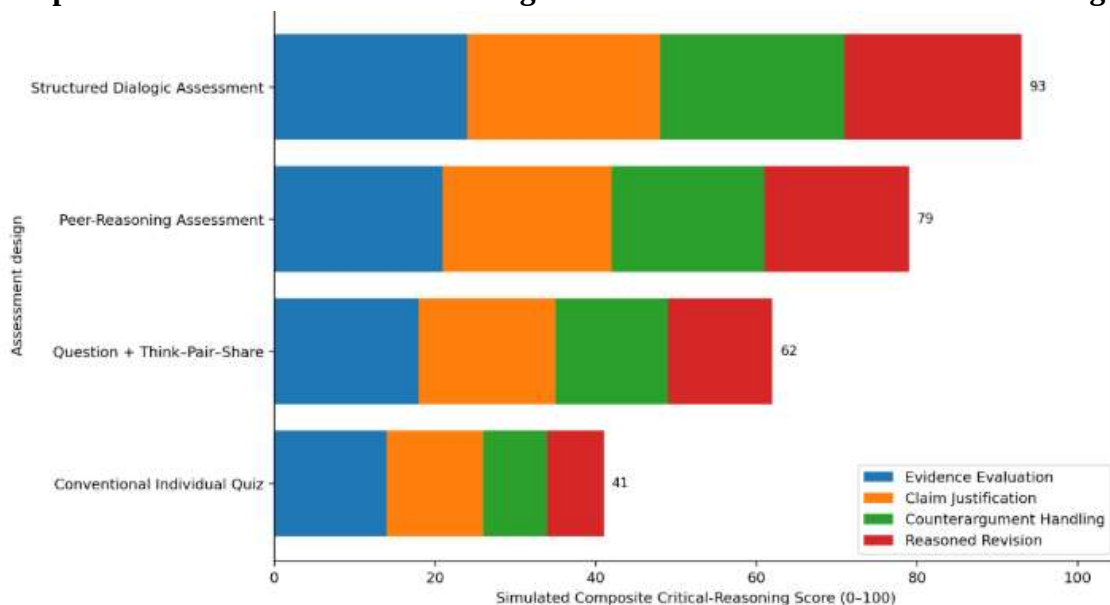
Assessment Design	Evidence Evaluation	Claim Justification	Counterargument Handling	Reasoned Revision	Composite CRDI
Share					
Peer-Reasoning Assessment	21	21	19	18	79
Structured Dialogic Assessment	24	24	23	22	93

Note: All numerical values are simulated for methodological illustration and do not represent empirical student outcomes.

The conventional quiz receives the lowest simulated score because it can evaluate individual judgments efficiently but provides limited direct evidence concerning how students react to opposing evidence or whether they can revise reasoning after criticism. This does not imply that individual assessment is educationally weak; independent assessment remains essential for establishing personal achievement. The limitation concerns the range of reasoning processes it makes observable.

5.2 Stacked-Bar Analysis of Critical-Reasoning Components

Graph 1: Simulated Critical-Reasoning Profile Across Scalable Assessment Designs



The stacked bar graph shows that the difference among the four assessment designs is not confined to one reasoning component. Evidence evaluation increases progressively, but the largest conceptual weakness of conventional individual quizzing appears in counterargument handling and reasoned



revision. These are precisely the reasoning activities that require students to encounter an alternative position and determine whether their original judgment remains defensible.

The graph also illustrates why dialogue needs structure. Think–pair–share produces improvement, but the peer-reasoning and structured-dialogic models generate stronger simulated gains because students are required to evaluate arguments rather than simply exchange answers.

The result is conceptually consistent with research showing that regulation scripts can strengthen the quality of dialogic peer assessment and that particular teacher dialogue moves can draw out critical thinking. The pedagogical implication is that the assessment prompt should specify the intellectual move expected from students.

5.3 A Scalable Implementation Model for Large Classes

A practical large-class implementation can operate within repeated short cycles rather than converting every lecture into a lengthy seminar. The instructor first presents a conceptually difficult question for which distractors represent plausible misconceptions rather than arbitrary wrong answers. Students commit individually. Response distributions then inform the instructor whether dialogue is warranted. Students subsequently compare reasoning in pairs or triads. Rather than using the generic instruction “discuss your answer,” they receive a compact dialogic prompt such as: State your claim; identify your strongest evidence; explain one assumption; challenge your partner's reasoning; decide whether your original position should change.

A second response captures revision. Importantly, students should not be rewarded simply for changing their answer. A learner who retains the original position after evaluating a weak objection may demonstrate stronger critical reasoning than a learner who changes merely because a peer appeared confident.

The instructor can then display anonymized reasoning samples or summarize contrasting arguments. Large-class technologies permit response aggregation, while expert judgment remains necessary to establish which arguments are disciplinary sound. Peer Instruction research demonstrates the practicality of such polling–discussion–reconsideration cycles in large classes, and active-learning evidence indicates that structured student engagement can remain beneficial even at scale.

6. Discussion

6.1 Dialogic Assessment Makes Reasoning Processes Visible

The central contribution of dialogic assessment is that it changes what counts as assessment evidence. Traditional assessments frequently record a student's final response, whereas dialogic assessment captures part of the intellectual pathway connecting evidence, judgment, criticism, and revision. This distinction matters because a correct answer does not necessarily demonstrate strong reasoning and an initially incorrect answer does not necessarily indicate poor learning potential.

A student who recognizes a flawed assumption after peer challenge and reconstructs the argument has demonstrated a form of critical reasoning that a single-response item cannot capture. Cui and Teo's work supports this emphasis on visible reasoning by showing how classroom discourse moves can draw out critical thinking through elaboration rather than simple answer production.



The assessment implication is not that dialogue should replace individual examinations. High-stakes decisions still require defensible evidence of individual achievement. Dialogic assessment is strongest as a recurring formative or low-stakes component that develops the reasoning students will later demonstrate independently. It enables misconceptions and unsupported arguments to become visible before they are consolidated into final work. Dialogic formative-feedback research similarly conceptualizes feedback as an interactive process through which learners interpret and use information rather than merely receive comments.

A further benefit is metacognitive. Students must determine not simply whether their answer changed but *why* it changed. Reasoned revision therefore creates an important distinction between intellectual flexibility and social conformity. The objective is not to reward agreement but to reward justified reconsideration.

6.2 Large Classes Require Distributed Dialogue Rather Than Instructor-Dominated Dialogue

The principal scalability problem in dialogic education is often framed as an instructor-capacity problem: one academic cannot hold meaningful individual dialogue with several hundred students simultaneously. The proposed framework changes the unit of dialogue. Students become temporary reasoning partners and evaluators, while the instructor designs the epistemic conditions, monitors aggregated evidence, samples arguments, and provides disciplinary synthesis.

Evidence from Peer Instruction and peer assessment suggests that such distributed activity can produce learning benefits when the process is carefully structured. The large-class mathematics evidence reported by Vozzo and colleagues further demonstrates that peer discussion can be associated with gains that persist beyond the immediate question. This makes peer dialogue particularly valuable where instructor feedback is necessarily limited by class size.

Distribution, however, creates quality-control requirements. Students can reinforce misconceptions, defer to dominant peers, or provide shallow feedback. Dialogic assessment therefore requires structured criteria and instructor visibility. Cheng and colleagues' findings concerning regulation scripts are particularly relevant because they show that peer assessment benefits from prompts that structure how feedback dialogue occurs.

Participation equity must also be considered. Large-class dialogue can reproduce social differences if confident speakers consistently dominate discussion. Individual commitment before peer exchange helps protect quieter students' initial reasoning, while rotating roles, written reasoning prompts, anonymous response systems, and brief pair interactions can widen participation. Active-learning research has shown that implementation intensity and design can influence equity outcomes, emphasizing that interaction should be intentionally structured rather than assumed to be automatically inclusive.

6.3 Assessment Quality Depends on the Design of the Dialogue Prompt

The success of dialogic assessment ultimately depends on what students are asked to do. A question with one obvious recall answer generates little intellectual benefit from discussion. Strong dialogic assessment items contain sufficient conceptual tension for reasonable students to initially disagree or to reach the same conclusion for different reasons.



Effective prompts should therefore ask students to identify the most relevant evidence, evaluate competing explanations, expose assumptions, compare causal claims, judge source credibility, identify counterexamples, or determine what additional evidence would alter a conclusion. This position is consistent with argumentation theory, which treats critical reasoning as an inherently dialectical process involving claims, supporting reasons, objections, and responses.

Feedback should likewise target reasoning rather than only correctness. An instructor can say that an answer is wrong, but a more powerful formative intervention identifies why the reasoning failed: the evidence may be irrelevant, the inference too broad, the causal explanation unsupported, the counterexample ignored, or the assumption unjustified.

Dialogic assessment therefore alters the instructor's role from sole respondent to designer and curator of reasoning encounters. In large classes, this shift is particularly valuable because the instructor's scarce attention can be directed toward misconceptions and argument patterns detected across the group rather than divided into hundreds of repetitive one-to-one explanations.

7. Conclusion

Dialogic assessment offers a promising means of reconciling two objectives that are often treated as incompatible in higher education: large-scale assessment and the development of sophisticated critical reasoning. Large classes inevitably constrain individualized instructor dialogue, but they do not eliminate students' capacity to reason with one another. When assessment is deliberately structured around individual commitment, evidence, justification, counterargument, peer challenge, and reasoned revision, dialogue can become a source of both learning and assessment evidence.

The simulation developed in this study illustrates the conceptual difference among four scalable assessment configurations. The conventional individual quiz achieved a simulated CRDI of 41, think-pair-share 62, peer-reasoning assessment 79, and structured dialogic assessment 93. These values are not empirical effects and should not be interpreted as predicted classroom gains. Their purpose is to demonstrate that richer assessment architectures can make a wider range of reasoning practices observable. The most important gain occurs not simply because students communicate more but because the assessment progressively requires them to evaluate evidence, justify claims, confront counterarguments, and explain revision.

The proposed model should therefore complement rather than replace individual assessment. Universities still require independent evidence that students can reason without peer assistance. Dialogic assessment is most valuable as a recurring formative and low-stakes architecture that gives students repeated practice in the intellectual moves later expected in essays, examinations, research, professional judgment, and problem solving. Existing evidence from dialogic teaching, peer assessment, Peer Instruction, and active learning provides a credible foundation for such an approach.

Future empirical research should test the proposed framework across disciplines and class sizes using validated critical-reasoning measures, performance tasks, dialogue transcripts, participation indicators, and longitudinal learning outcomes. Particular attention should be given to whether reasoned revision transfers to independent assessment, whether peer dialogue benefits quieter and academically less confident students, and which combinations of digital polling and human facilitation preserve both scalability and assessment validity. The central conclusion is that dialogue strengthens assessment when



it requires students to expose the reasons behind judgments and makes those reasons available for challenge, comparison, and revision. Large classes change how such dialogue must be organized, but they need not remove it from serious academic assessment.

References

1. Brindley, S., & Marshall, B.-J. (2014). "Resisting the rage for certainty": Dialogic assessment: A case study of one secondary English subject classroom in the UK. *Assessment in Education: Principles, Policy & Practice*, 21(4), 405–424.
2. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
3. Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. <https://doi.org/10.1007/BF00117714>
4. Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
5. Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57–66. <https://doi.org/10.1080/14703290601081332>
6. Keppell, M., & Carless, D. (2006). Learning-oriented assessment: A technology-based case study. *Assessment in Education: Principles, Policy & Practice*, 13(2), 179–191. <https://doi.org/10.1080/09695940600703944>
7. Espasa, A., Guasch, T., Mayordomo, R. M., Martínez-Melo, M., & Carless, D. (2018). A Dialogic Feedback Index measuring key aspects of feedback processes in online learning environments. *Higher Education Research & Development*, 37(3), 499–513. <https://doi.org/10.1080/07294360.2018.1430125>
8. Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52(1), 17–37.
9. Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, 69, 963–976.
10. Henderson, M., Phillips, M., & Ryan, T. (2018). Designing for technology-enabled dialogic feedback. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing Evaluative Judgement in Higher Education: Assessment for Knowing and Producing Quality Work* (pp. 117–126). Routledge.
11. Adie, L. (2018). The development and application of coding frameworks to explore dialogic feedback interactions and self-regulated learning. *British Educational Research Journal*, 44(4), 704–723. <https://doi.org/10.1002/berj.3463>
12. Hill, J., & West, H. (2019). Improving the student learning experience through dialogic feed-forward assessment. *Assessment & Evaluation in Higher Education*, 45(1), 82–97. <https://doi.org/10.1080/02602938.2019.1608908>
13. Tam, A. C. F. (2020). Undergraduate students' perceptions of and responses to exemplar-based dialogic feedback. *Assessment & Evaluation in Higher Education*, 46(2), 269–285.



- <https://doi.org/10.1080/02602938.2020.1772957>. The study examined how students responded to dialogic feedback and identified the importance of teacher scaffolding and supportive environments.
14. Rao, N., & Norton, L. (2020). Dialogic feedback: What we know and what remains to be known. In *On Your Marks: Learner-focused Feedback Practices and Feedback Literacy* (pp. 160–168). Advance HE.
 15. Macleod, G., et al. (2020). Balancing supportive relationships and developing independence: An activity theory approach to understanding feedback in context for Master's students. *Assessment & Evaluation in Higher Education*, 45(7), 958–972. <https://doi.org/10.1080/02602938.2020.1719976>
 16. Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, 85(4), 475–511.
 17. Alexander, R. J. (2018). Developing dialogic teaching: Genesis, process, trial. *Research Papers in Education*, 33(5), 561–598. <https://doi.org/10.1080/02671522.2018.1481140>. The intervention focused on using classroom talk to improve student engagement and learning.
 18. Segal, A., Snell, J., & Lefstein, A. (2017). Dialogic teaching to the high-stakes standardised test? *Research Papers in Education*, 32(5), 596–610. <https://doi.org/10.1080/02671522.2016.1225803>.
 19. Chen, P., Tolmie, A. K., & Wang, H. (2017). Growing the critical thinking of schoolchildren in Taiwan using the Analects of Confucius. *International Journal of Educational Research*, 84, 43–54. <https://doi.org/10.1016/j.ijer.2017.02.002>. The study provides evidence connecting dialogic teaching with development of exploratory thinking.
 20. Teo, P. (2019). Teaching for the 21st century: A case for dialogic pedagogy. *Learning, Culture and Social Interaction*, 21, 170–178. <https://doi.org/10.1016/j.lcsi.2019.03.009>. The paper specifically discusses dialogic pedagogy in relation to communication, collaboration and critical thinking.