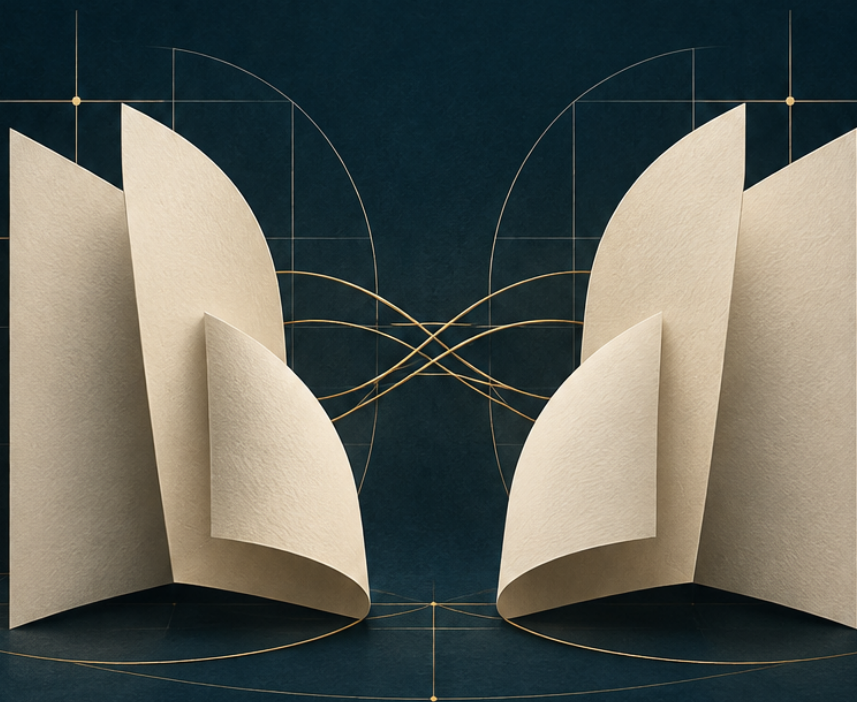


BEYOND THE RUBRIC

Reconceptualizing Assessment
as Dialogue



DR. NAEEM FATIMA

LUMINA LITERATI PUBLISHING

LUMINA LITERATI PUBLISHING

Beyond the Rubric

Reconceptualizing Assessment as Dialogue



Dr. Naeem Fatima

Lumina Literati Publishing

2026

Beyond the Rubric

Reconceptualizing Assessment as Dialogue

© 2026 Dr. Naeem Fatima.

The author retains 100% copyright ownership.

Published in 2026 by Lumina Literati Publishing.

luminalliterati.com

ISBN 978-627-7813-61-1

DOI 10.65758/isbn.9786277813611

Open access · CC BY-NC-ND 4.0

This work is licensed under the Creative Commons

Attribution-NonCommercial-NoDerivatives 4.0 International license.

Noncommercial copying and sharing are permitted with appropriate attribution, preservation of copyright and license notices, and a link to the license. Adapted versions may not be distributed under this license. No additional restrictions may be imposed on uses the license permits.

creativecommons.org/licenses/by-nc-nd/4.0/

Uses permitted by copyright exceptions and limitations remain unaffected. For uses beyond the license and those exceptions, obtain permission from the copyright holder.

Cover artwork created by Lumina Literati Graphics.

Set in EB Garamond and Cormorant Garamond.

Contents

About the author	v
1. Foundations and Framing	1
1.1 From Measurement to Meaning: The Assessment Paradigm and Its Discontents	1
1.2 The Reconsidered Rubric: Affordances, Seductions, and the Costs of Criteria Compliance	10
1.3 Dialogue as a Theoretical Lens: Bakhtin, Vygotsky, and the Social Construction of Judgement	20
1.4 Scope, Aims, and Guiding Questions	30
2. The Architecture of Dialogic Assessment	36
2.1 Defining Dialogic Assessment: Core Principles and Boundary Conditions	36
2.2 Feedback as Dialogue, Not Delivery: From Information to Interaction	46
2.3 Co-Constructing Criteria and Cultivating Evaluative Judgement	55
2.4 Designing for Dialogue: Conditions, Constraints, and Trade-offs	63
3. Dialogue in Practice	71
3.1 The Feedback Conversation: Dialogic Feedback in Disciplinary Writing	71
3.2 Co-Constructed and Negotiated Assessment: Students as Partners in Judgement	80
3.3 Patchwork, Portfolios, and Programmatic Assessment: Dialogue Across Time	89
3.4 Peer Dialogue and Calibration: Assessment as Distributed Conversation	98

4. Mechanisms and Dynamics	107
4.1 Trust, Power, and Vulnerability: The Relational Engine of Assessment Dialogue	107
4.2 From Telling to Internalising: How Dialogue Builds Evaluative Judgement and Self-Regulation	115
4.3 Equity, Inclusion, and the Politics of Voice in Assessment Dialogue	124
4.4 An Integrated Model: The Dialogic Assessment Framework (DAF)	132
5. Beyond the Rubric in Practice	142
5.1 Redesigning Curriculum and Assessment Regimes for Dialogue	142
5.2 Faculty Development and Assessment Literacy: Preparing Educators for Dialogue	151
5.3 Quality, Standards, and Policy: Reconciling Dialogue with Accountability	159
5.4 Generative AI and the Future of Assessment Dialogue	166
5.5 Epilogue: Assessment as a Practice of Recognition	174
References	179

About the author

Dr. Naeem Fatima is a distinguished educator, researcher, and author with a career spanning over twenty-five years in higher education. She currently serves as the Principal of Fazaia College of Education for Women, Peshawar, where she provides academic and administrative leadership while fostering a transformative learning environment.

Her extensive teaching portfolio encompasses a sophisticated range of subjects, including English Language Teaching, Technical Writing, Functional English, and Research Methodology. In the domain of teacher education, she has profoundly influenced future educators through her courses in Educational Psychology and Teaching Methodology.

As a prolific scholar, Dr. Fatima has authored several influential books that bridge theory and practice. Her publications include *Attention Modulation in Reading*, *Transformative Pedagogies in Higher Education*, *Theoretical Crossroads: Integrating Formal and Cognitive Approaches to Language*, and *Use of AI in Social Sciences Research*. These works reflect her interdisciplinary approach and her commitment to advancing educational and linguistic thought.

Her research interests are both deep and expansive, firmly rooted in Systemic Functional Linguistics and extending into dynamic, socially relevant applications. Her investigative work includes Interaction Dynamics, Feedback Mechanisms, and Innovative Language Pedagogies, with a particular focus on Translanguaging and Critical Discourse Analysis — exploring themes of language, power, and resistance. Her research portfolio further encompasses an intercultural study of compliments and compliment responses, examinations of the role of technology in teacher–student communication, and investigations into the impact of Project-Based Learning on students’ critical thinking, creativity, and problem-solving skills.

Through her leadership, pedagogy, and scholarship, Dr. Fatima continues to be a vital force in shaping educational practices and linguistic inquiry, dedicated to empowering both students and colleagues through knowledge, critical reflection, and innovative methodologies.



Foundations and Framing

The Limits of the Rubric and the Case for Dialogue

1.1 From Measurement to Meaning: The Assessment Paradigm and Its Discontents

The contemporary university is saturated with assessment. Students enter courses already aware that their intellectual work will be converted into grades, credits, classifications, grade-point averages, transcripts, progression decisions, scholarship eligibility, professional accreditation and, eventually, employability signals. Teachers design modules around learning outcomes, construct rubrics, moderate scripts, defend marks at examination boards and produce evidence for quality assurance. Institutions aggregate results into performance indicators, retention data, programme reviews and external accountability reports. Assessment, in this sense, has become one of the principal languages through which higher education describes itself. Yet this language is increasingly strained. The dominant paradigm of assessment-as-measurement has been remarkably successful in producing comparable scores, auditable records and defensible certification decisions. It has been less successful in sustaining educational relationships through which judgement becomes meaningful, feedback becomes usable and students learn to participate in the standards of a field.

This book begins with the claim that the limits of assessment-as-measurement are not merely technical. They cannot be solved simply by more detailed rubrics, more precise grade descriptors, more elaborate

moderation procedures or more efficient digital marking systems. These instruments may improve some aspects of transparency and consistency, but they do not address the deeper difficulty: assessment is not only a measurement event but a communicative act. It is a transaction of meaning between teachers, students, disciplines and institutions. When it is treated primarily as the one-way measurement of student performance against pre-specified criteria, it weakens the dialogic conditions under which students come to understand what quality means, how standards are interpreted and how judgement is formed. The failures of contemporary assessment therefore require attention to communication and relationships as well as technical design.

1. Three Purposes of Assessment: Of, For and As Learning

A useful starting point is the now-familiar distinction among assessment of learning, assessment for learning and assessment as learning. Assessment of learning refers to the summative and certificatory function of assessment. Its main concern is to determine what has been achieved after a period of instruction. It produces marks, grades, awards, classifications and credentials. In higher education, this is the function most directly connected to institutional legitimacy: universities must be able to say who has passed, who has failed, who has met professional standards and who is entitled to a degree.

Assessment for learning, by contrast, refers to the formative use of evidence to improve learning while learning is still taking place. In the tradition associated with Black and Wiliam (1998), assessment becomes formative when information about student understanding is used by teachers and learners to modify teaching and learning activities. The emphasis shifts from recording achievement to generating feedback loops: Where is the learner now? Where does the learner need to go? What action can close the gap? Assessment for learning therefore depends on interpretation, response and timing. Evidence has formative value only if it can be understood and acted upon before the learning opportunity has passed.

Assessment as learning goes further by locating assessment within the development of learner self-regulation. Here students are not simply recipients of teacher feedback but active participants in judging the quality of their own work. They learn to monitor performance, interpret criteria, compare exemplars, identify weaknesses and make decisions about improvement. This purpose is closely related to Boud's (2000) argument for sustainable assessment: higher education should prepare students not only to be assessed by others but to become capable judges of their own learning beyond the course, the module and the degree.

The measurement paradigm privileges assessment of learning because summative certification is the purpose most compatible with quantification, standardisation and audit. Grades travel more easily than conversations. Rubrics are easier to archive than dialogue. A percentage mark is more administratively useful than a student's developing capacity to make evaluative judgements. As a result, assessment for learning and assessment as learning are often subordinated to the summative system. Feedback is written after the grade has already done its work. Self-assessment is included as a minor reflective activity but rarely given structural authority. Formative tasks are valued by students only when they appear to predict, rehearse or secure the summative grade. The system may formally acknowledge three purposes, but its architecture repeatedly privileges the first.

2. The Historical Rise of the Measurement Model

The measurement model did not emerge accidentally. Its intellectual roots lie partly in psychometrics and the rise of standardized testing, where the central problems were reliability, comparability, validity, scaling and selection. In this tradition, assessment is imagined as an instrument designed to produce an accurate estimate of an underlying trait, ability or level of achievement. The ideal assessment is one in which irrelevant variation is minimized, marker subjectivity is controlled and results can be compared across individuals and contexts.

This model acquired particular force in mass higher education. As universities expanded, assessment had to manage larger cohorts, more diverse student populations and greater public scrutiny. The small tutorial model, in which standards were transmitted through sustained apprenticeship, became increasingly difficult to maintain. Modular degree structures intensified this development. Learning was divided into discrete units, each with specified outcomes, credit values, assessment tasks and grade weightings. The module became the primary administrative container of learning, and the assessment task became the point at which the module demonstrated that its outcomes had been achieved.

Outcomes-based education reinforced this logic. By requiring programmes to define intended learning outcomes and align teaching, learning activities and assessment, it promised transparency and coherence. In principle, this alignment can be educationally valuable. It can clarify expectations, reduce hidden rules and make assessment more purposeful. In practice, however, it has often encouraged a documentary culture in which learning must be translated into measurable outcomes, outcomes into criteria and criteria into auditable evidence. What cannot be easily specified becomes harder to defend. What cannot be defended becomes institutionally vulnerable.

The audit culture of quality assurance further strengthened assessment-as-measurement. External examiners, accreditation bodies, programme reviews and regulatory agencies all require evidence that standards are being maintained. Marks, rubrics, moderation forms, grade distributions and outcome maps function as signs of control. Power's (1997) analysis of the audit society and Strathern's (2000) account of audit cultures help explain why such systems expand: they promise trust through verification, but in doing so they can displace professional trust with documentary proof. The result is not simply more assessment but a new relationship to assessment. Assessment becomes evidence that the institution is managing risk.

Torrance's (2007) critique of post-secondary education and training is especially important here. He argues that the increased explicitness of learning objectives, assessment criteria and feedback can produce a paradoxical effect. Practices initially designed to support learning may come to dominate learning. Students learn to perform the visible requirements of the assessment system rather than engage deeply with the knowledge, problems and uncertainties of the discipline. Assessment for learning is thereby converted into "assessment as learning" in a negative sense: assessment procedures become the curriculum that students actually study.

3. Core Discontents of the Measurement Paradigm

The discontent of this system is now visible across higher education. One is concern about grade inflation. Rising marks alone do not establish declining standards; changes in attainment, assessment and student populations also require consideration. When grades serve as the dominant currency of academic value, pressure grows on both sides of the assessment relationship. Students experience grades as high-stakes signals of future opportunity; teachers experience grading as a site of complaint, negotiation and institutional scrutiny. The grade becomes both judgement and commodity. In such conditions, the upward drift of marks is not simply a technical problem of standard-setting. It may also involve the meanings attached to grades in competitive educational economies.

Grade anxiety is the other side of the same process. Students often approach assessment less as an opportunity to test understanding than as a threat to identity, progression and prospects. The question "Will this be on the exam?" is not evidence of intellectual laziness; it is a rational response to a system that concentrates value in summative outcomes. In the modular exam economy, students may calculate effort according to credit weightings, deadlines and grade returns. A ten-credit module due next week may command more attention than a difficult concept central to disciplinary formation. Learning is scheduled around assessment pressure rather than intellectual development.

A second discontent is the rise of strategic criteria compliance. Rubrics and criteria were introduced partly to make assessment fairer and more transparent. Yet when criteria are treated as exhaustive descriptions of quality, students may learn to comply rather than inquire. The undergraduate essay offers a familiar example. A student writing on social theory may ask not, “What is the strongest interpretation of this argument?” but “How many theorists do I need to mention for a first?” The essay becomes a checklist: define key terms, include critical analysis, use sources, link to the question, reference correctly. These are not unimportant matters. The problem arises when the criteria become a substitute for judgement. Students learn to display the signs of academic performance without necessarily entering the deeper conversation that the essay was meant to invite.

A third discontent is the marginalisation of feedback. Feedback is widely celebrated in policy language, yet in practice it is often structurally weak. It arrives too late, attached to a grade that absorbs the student’s attention. It is written as justification rather than dialogue. It explains why a mark was awarded but does not necessarily create conditions for revision, response or future judgement. Students may read comments selectively, looking for clues to the next grade rather than engaging with the intellectual substance of the response. Teachers, in turn, may experience feedback as labour that is neither read nor rewarded. A communicative act becomes a bureaucratic residue of marking.

A fourth discontent is the potential tension between reliability and validity when assessment is narrowed to secure consistency. Standardisation can support dependable judgements, but it does not by itself establish whether an assessment represents the learning that matters. Multiple-choice tests can assess complex reasoning when well designed, while essays, projects, portfolios and performances can offer evidence of capabilities that selected-response tests capture less directly. Neither format guarantees validity, and reliability and validity are not inherently opposed. The design challenge is to obtain sufficiently dependable

evidence of the intended learning, using appropriate tasks, criteria and moderation. This book treats that challenge also as a communicative problem: complex academic judgement cannot be fully externalised into rules because standards are interpreted within communities of practice.

Finally, there is the erosion of trust between students and teachers. In a measurement-dominated system, the teacher is easily positioned as gatekeeper and the student as claimant. Students ask, “What do you want in this assignment?” Teachers respond by clarifying expectations, but the exchange already reveals a damaged relationship. The assignment is no longer understood primarily as an invitation into a problem, question or practice. It becomes an attempt to infer the assessor’s preferences. The student seeks security; the teacher seeks independence. Both are acting rationally within a system that has made the grade more visible than the conversation.

4. Three Higher Education Scenes

Consider first the marking of essays against criteria. A rubric may distinguish argument, evidence, structure, originality and referencing. Used well, it can support shared attention to important features of academic work. Used poorly, it fragments quality into separable boxes. The marker knows that the difference between a competent and an excellent essay often lies not in the isolated presence of each criterion but in the relation among them: the way evidence is selected because the argument demands it; the way structure reflects conceptual control; the way originality emerges from disciplined engagement with sources. Such judgement is holistic, interpretive and disciplinary. A rubric can point toward it, but it cannot replace the conversation through which students learn to recognize it.

Consider second the modular exam economy. In many degree programmes, students move from module to module, deadline to deadline, accumulating credits. Examinations and assignments are distributed across the semester like economic demands on attention.

Students quickly learn the exchange rate: how much effort is required, what each task is worth and which forms of preparation produce marks most efficiently. This is not irrational behaviour; it is an adaptation to the assessment ecology. Yet this ecology may produce surface or strategic learning even when teachers intend depth. Students may revise for the exam, perform adequately and then abandon the knowledge because the next module has already begun. The system certifies achievement but does not necessarily secure durable understanding.

Consider third the common exchange: “What do you want in this assignment?” On the surface, this is a request for clarification. At a deeper level, it expresses the student’s dependence on hidden standards and the teacher’s responsibility to make expectations intelligible. The question becomes problematic only when it replaces the disciplinary question. In a dialogic assessment culture, the teacher might respond by discussing the purpose of the task, showing examples of stronger and weaker work, inviting students to evaluate drafts and helping them internalize standards. In a measurement culture, the response is more likely to be a list of requirements. The student receives instructions but not necessarily induction into judgement.

5. Assessment as Communicative Act

The central claim of this book is that assessment is irreducibly communicative. It always says something: about what counts as knowledge, what forms of performance matter, what kinds of persons students are becoming and what relationships exist between teachers and learners. A grade says, “This is where your work has been placed.” A comment says, “This is how your work has been interpreted.” A rubric says, “These are the visible categories through which quality will be discussed.” A moderation process says, “This judgement must be defensible within a community.” Even silence communicates: when feedback is absent, delayed or unusable, students learn that assessment is primarily about ranking rather than learning.

To call assessment communicative is not to reject measurement altogether. Universities must still certify achievement, protect standards and make consequential decisions. Nor is it to romanticize dialogue as if all assessment could be converted into conversation. Rather, it is to argue that measurement is a partial and dependent function. Its legitimacy depends on prior and continuing acts of communication: the negotiation of standards, the interpretation of criteria, the modelling of quality, the uptake of feedback and the cultivation of student judgement. When these communicative conditions collapse, measurement becomes hollow. It may produce numbers, but the numbers no longer carry the educational meaning they claim to represent.

This is why measurement alone has limits as an educational approach. It can improve consistency but not necessarily understanding. It can generate evidence but not necessarily trust. It can classify performance but not necessarily develop judgement. It can certify learning while simultaneously reshaping learning into strategic compliance. The problem is not that assessment measures too much, but that it too often measures without sufficient dialogue about what is being measured, why it matters and how students can come to participate in the standards by which they are judged.

The reconstructive task, therefore, is not to abandon assessment but to reconceptualize it. Assessment must be redesigned as a dialogic practice in which summative judgement, formative feedback and self-regulatory learning are held together rather than separated into competing functions. Such a reconstruction requires attention to relationships, temporality and meaning. It asks how students encounter standards before the final submission, how feedback becomes usable, how teachers' professional judgement is made discussable without being reduced to mechanical criteria and how institutions can sustain public trust without turning all trust into audit.

The chapters that follow develop this argument in greater detail. They examine rubrics, feedback, moderation, student agency, teacher

judgement and institutional accountability through the lens of assessment as dialogue. The aim is not to deny the need for reliable certification, but to show that certification without communication is educationally impoverished. Assessment can only support learning when it is understood not as the endpoint of teaching but as one of the central media through which teaching and learning are related. Beyond the rubric lies not the absence of standards, but a richer account of how standards become meaningful between people.

1.2 The Reconsidered Rubric: Affordances, Seductions, and the Costs of Criteria Compliance

If assessment-as-measurement has a characteristic object, it is the rubric. The rubric is not merely a marking aid. It is the emblematic technology of contemporary assessment because it condenses many of the system's governing aspirations: transparency, consistency, accountability, defensibility and efficiency. In a single grid, the rubric promises to translate academic judgement into visible criteria, graded descriptors and levels of performance. It appears to offer protection against arbitrariness for students, support for novice learners, guidance for markers, evidence for external examiners and documentation for institutional quality assurance. It is therefore unsurprising that rubrics have become central to higher education assessment design.

Yet the rubric's very attractiveness is also the source of its danger. It offers a solution to real problems, but it does so by representing those problems in a particular way. It assumes that academic standards can be named in advance, divided into criteria, described across levels and applied to work once submitted. This assumption is not simply false. Many features of academic work can and should be made explicit. The difficulty is that the rubric tends to exaggerate the codifiability of judgement. It makes assessment look more stable, more transparent and more rule-governed than it actually is. The issue, therefore, is not whether rubrics are

good or bad. The issue is what kind of educational relation the rubric constructs between teacher, student, standard and work.

1. Legitimate Affordances: Why Rubrics Became So Attractive

Any serious critique of rubrics must begin by acknowledging their legitimate affordances. Rubrics emerged partly because students had good reasons to distrust vague assessment practices. In many higher education settings, students encounter assignments with limited explanation of what counts as quality. They may receive comments such as “needs more critical analysis,” “argument lacks depth,” or “good structure,” without understanding what these judgements mean in practice. Rubrics respond to this problem by naming dimensions of performance: argument, evidence, originality, coherence, method, presentation, referencing, analysis or reflection. In doing so, they can reduce the opacity of assessment.

The first affordance, then, is transparency. A rubric can make visible the aspects of work to which assessors are expected to attend. For novice students, especially those entering unfamiliar disciplinary cultures, this is not trivial. A first-year student asked to write a sociology essay may not yet understand what distinguishes description from analysis, assertion from argument, or opinion from evidence-based interpretation. A rubric can provide an initial map of the terrain. It tells students that academic quality is not a mysterious personal preference of the teacher but involves recognizable features of performance.

A second affordance is consistency across markers. In large modules, where dozens or hundreds of scripts are divided among teaching assistants, tutors and lecturers, rubrics can support shared expectations. They provide a common language for moderation and reduce the likelihood that one marker rewards structure while another rewards originality, or that one marker penalizes minor referencing errors more severely than another. This matters not because perfect consistency is

possible, but because students are entitled to expect that their work will not be judged according to wholly private standards.

A third affordance is the reduction of arbitrariness. The rubric disciplines the assessor as well as the student. It requires teachers to articulate what they are looking for, to connect marks to criteria and to justify decisions in ways that can be reviewed. This can be particularly important in programmes where assessment has historically depended on tacit departmental norms. Rubrics can compel teams to discuss standards that might otherwise remain implicit. In this respect, they do not merely record standards; they can provoke useful professional conversations among assessors.

A fourth affordance is scaffolding. Reviews by Reddy and Andrade (2010) and Panadero and Jonsson (2013) suggest that rubrics can support student performance, self-regulatory activity and awareness of expectations, although findings vary with their design and use. Their formative potential is better realised when students are taught how to use them, compare them with exemplars, apply them to drafts and discuss their interpretation.

A fifth affordance is defensibility under appeal. In the contemporary university, assessment decisions must increasingly be explainable. Students may challenge marks, external examiners may question grade distributions and institutions may need to demonstrate procedural fairness. A rubric provides a documentary basis for the judgement. It shows that the assessment was not wholly impressionistic. It offers evidence that the mark was connected to pre-established criteria. In this administrative sense, rubrics are not simply pedagogical tools; they are instruments of institutional risk management.

These affordances explain the rubric's appeal. It responds to real anxieties: students' anxiety about hidden expectations, teachers' anxiety about inconsistency, institutions' anxiety about appeals and regulators' anxiety about standards. A university without any explicit criteria would

be difficult to defend. The problem is that the rubric's legitimate usefulness can lead to an inflated belief in its power.

2. The Seduction of Transparency

The central seduction of the rubric is transparency. It promises to make tacit standards explicit. In doing so, it addresses a genuine ethical problem in higher education: students should not be assessed according to rules they have had no opportunity to learn. The work of Rust, O'Donovan and Price (2005) is important because it shows both the necessity and the insufficiency of explicitness. Their argument is not that standards should remain hidden, but that students come to understand standards through a combination of explicit explanation, exemplars, dialogue and participation in assessment cultures. Standards are not simply transmitted through written criteria. They are learned through use.

Sadler's (2010) work sharpens this point. Complex academic standards involve tacit evaluative knowledge. In disciplines, experts often recognize quality before they can fully define it. A historian may recognize a subtle use of evidence; a literary scholar may recognize interpretive elegance; a sociologist may recognize theoretical control; a designer may recognize coherence between concept and form. These judgements are not arbitrary, but neither are they fully reducible to explicit rules. They depend on accumulated experience, comparison with prior examples and immersion in disciplinary practice.

The rubric attempts to codify this knowledge by converting it into descriptors. For example, a top-band descriptor for an essay might state: "Demonstrates excellent critical analysis, sustained argument, independent engagement with literature and clear structure." Such language is useful, but it does not settle the matter. What counts as "excellent"? How much independence is enough? When does structure become mechanical? When does criticality become superficial disagreement? Students may read the words without possessing the

interpretive background needed to apply them. The transparency is therefore partial.

This partiality is not a failure of wording alone. It is a limit of language in relation to complex performance. Some standards can be specified clearly: word limits, citation formats, inclusion of required sections, methodological transparency, ethical approval. Other standards are more relational and qualitative: elegance, insight, depth, coherence, originality, persuasiveness, intellectual risk. These cannot be fully defined in advance because their meaning depends on the particular work being judged. A brilliant essay may violate an expected structure in order to produce a more powerful argument. A technically compliant essay may satisfy every criterion and still lack intellectual force.

The seduction of transparency lies in the belief that more detailed descriptors will solve this problem. If students do not understand “critical analysis,” the rubric adds sub-descriptors. If markers disagree, the rubric adds levels. If appeals increase, the rubric becomes more specific. Yet increasing specificity can create a false clarity. It may give the impression that quality has been captured when in fact only its surface indicators have been listed. The more elaborate the rubric becomes, the more it risks mistaking description for understanding.

3. Unintended Consequences: Compliance, Fragmentation and Narrowing

The first major unintended consequence is criteria compliance. Torrance’s (2007) critique of “assessment as learning” is relevant here: explicit objectives, criteria and feedback, originally intended to support learning, can come to dominate learning. Students learn how to satisfy the assessment apparatus rather than engage with the underlying knowledge. The question shifts from “What is worth understanding?” to “What does the rubric require?”

This is visible in the common essay-planning strategy of working backwards from the rubric. A student may allocate paragraphs to criteria:

one paragraph for theory, one for evidence, one for critique, one for application. The resulting essay may appear well aligned with expectations but remain intellectually thin. It performs the signs of academic work without necessarily developing an argument. The student has learned compliance, not judgement.

The second consequence is fragmentation. Analytic rubrics divide performance into parts. This can help identify strengths and weaknesses, but it can also distort the nature of complex achievement. In many academic tasks, quality lies in the integration of elements rather than their separate presence. Argument and evidence are not independent. Structure is not independent of meaning. Originality is not independent of knowledge. Criticality is not a detachable skill. When a rubric treats these as separate boxes, it may encourage students and markers to think atomistically.

This fragmentation affects holistic judgement. Experienced assessors often judge complex work by apprehending the relation among parts. They ask whether the work holds together, whether the claims are warranted, whether the student has understood the problem, whether the performance has intellectual life. A rubric can support such judgement, but it cannot replace it. When the rubric becomes the judgement rather than a guide to judgement, it risks reducing the assessor's professional discernment to a checklist.

The third consequence is curricular narrowing. Rubrics privilege what can be specified. Over time, assessment tasks may be designed around criteria that are easy to describe and defend. Complex, emergent, risky or creative forms of learning may be marginalized because they are harder to rubricize. The curriculum subtly contracts around what can be measured, evidenced and appealed. Teachers may avoid tasks whose outcomes are uncertain, even when such uncertainty is educationally valuable.

The fourth consequence is false certainty. Students may believe that if they follow the rubric, they are entitled to a high mark. This belief is understandable. The rubric appears contractual: do these things and

receive this grade. But academic judgement is rarely so mechanical. A student may include all required elements yet fail to produce a persuasive whole. Conversely, a student may exceed the rubric's language by doing something unexpected and powerful. When students interpret rubrics as guarantees, disappointment becomes likely. The appeal then becomes not only about the grade but about the broken promise of transparency.

4. The Reliability Paradox

Rubrics are often defended because they improve reliability. This claim is partly justified. A well-designed rubric can improve inter-marker agreement, especially where assessment concerns observable features or where markers receive calibration training. It can reduce extreme variation and provide a shared vocabulary for moderation. In large-scale assessment, this is valuable.

The paradox is that a rubric may improve agreement on form while doing little to secure the assessment of substance. Markers may agree that a student has included references, followed structure, used appropriate terminology and addressed required sections. They may still disagree about the depth, originality or intellectual significance of the work. In fact, the more complex the task is, the more judgement depends on features that resist easy codification.

This produces reliability-validity tension. Rubrics can make marking more consistent by narrowing attention to specified criteria. But if the criteria capture only the most visible aspects of performance, increased agreement may be achieved at the expense of validity. Assessors may become more consistent in judging what is easiest to describe rather than what is most educationally important. In this sense, reliability can become a formal achievement that conceals substantive loss.

The assessment of originality illustrates the problem. A rubric may include originality as a criterion, but originality is context-dependent. In a first-year essay, originality may mean making a thoughtful connection between two readings. In a doctoral thesis, it may mean producing a new

contribution to knowledge. In creative practice, it may involve risk, transformation or reconfiguration of materials. The same word cannot carry the same meaning across contexts without interpretive work. A rubric can name originality; it cannot fully determine its presence.

Similarly, in research-based assignments, the most significant achievement may emerge from the way a student frames the problem. Yet problem-framing is difficult to separate from method, evidence and argument. A rubric may allocate marks to each, but the intellectual value lies in their configuration. This is why experienced assessors often resist purely additive marking. They know that excellence is not the sum of separately adequate parts.

The reliability paradox therefore reveals the limits of the rubric as a measurement device. It can stabilize judgement through shared categories, but those categories inevitably simplify some aspects of performance. Some simplification is necessary. No assessment system can operate without categories. But when simplification is mistaken for truth, assessment begins to misrecognize the achievement it claims to represent.

5. Rubric as Monologue

The deeper conceptual problem arises when a rubric is used as a monologue. It speaks before work exists. It specifies, classifies and anticipates. It is handed down from teacher, programme or institution to student as a fixed representation of expected quality. When simply distributed and applied without discussion, it remains one-directional. It does not listen to the students' emerging argument. It cannot respond to an unexpected insight. It cannot revise its categories in light of a performance that redefines the task.

This matters because complex achievement is emergent. In serious academic work, the quality of a performance often becomes visible only through the work itself. A student may approach a question in an unexpected way, draw on unfamiliar sources, challenge the assumptions of the task or synthesize materials in a manner not anticipated by the

rubric. The assessor must then ask not merely whether the work matches the criteria, but whether the criteria are adequate to the work. This is not an invitation to arbitrariness. It is recognition that judgement is situated. Interpretive flexibility must still respect published requirements and equitable treatment; it is not a licence to introduce undisclosed grading criteria.

Rubrics struggle with emergence because they are designed to reduce uncertainty before the event. Their administrative value depends on stability. But educational judgement often requires responsiveness. A teacher assessing a reflective portfolio, for example, may need to consider the relation between experience, interpretation, self-critique and professional growth. These relations cannot be fully specified in advance because they depend on the student's trajectory. A rubric may identify relevant dimensions, but the meaning of those dimensions must be negotiated through the work.

This is why rubrics can become alienating. Students receive the rubric as an external demand. Teachers apply it as an institutional requirement. Both may comply with it without experiencing it as a living account of quality. The rubric then stands between teacher and student rather than mediating a conversation between them. It becomes the voice of assessment without the presence of dialogue.

6. From Transmission to Dialogue

If standards are partly tacit, socially situated and context-dependent, then sharing them cannot be understood as simple transmission. A rubric can transmit words, categories and levels. It cannot by itself transmit judgement. Students learn standards by seeing them enacted, comparing examples, attempting work, receiving response, revising, discussing alternatives and gradually internalizing criteria of quality. This is a dialogic process.

The implication is not that rubrics should be abandoned. Rather, they must be repositioned. A rubric should be treated as a provisional artefact

within a wider ecology of assessment dialogue. It can open conversation, but it should not close it. It can name possible dimensions of quality, but it should not pretend to exhaust them. It can support fairness, but it should not replace professional judgement. It can guide students, but it should not train them merely to comply.

A dialogic approach would ask different questions of the rubric. Not only: Are the criteria clear? But also: How will students learn what these criteria mean? Not only: Do markers apply the rubric consistently? But also: What conversations enable shared interpretation? Not only: Does the rubric defend the grade? But also: Does it help students develop evaluative judgement? Not only: Does it make assessment transparent? But also: What remains necessarily tacit, and how will students be invited into that tacit knowledge?

The pivot to dialogue begins here. The problem with the rubric is not that it says too much, but that it speaks alone. Standards become meaningful when they are discussed, exemplified, contested and applied in context. Assessment must therefore move beyond the fantasy of complete specification toward practices that make judgement shareable. The rubric may remain, but it must no longer be treated as the final authority. It should become one participant in a broader conversation about quality, learning and disciplinary belonging.

1.3 Dialogue as a Theoretical Lens: Bakhtin, Vygotsky, and the Social Construction of Judgement

The preceding sections have argued that measurement alone has limits as an educational approach and that the rubric, while valuable, cannot bear the full weight placed upon it. The present section establishes the theoretical basis for the book's alternative: assessment as dialogue. This claim is not intended metaphorically. "Dialogue" does not mean simply that teachers should be friendlier, that students should be allowed to ask questions, or that feedback should be written in a conversational tone. It names a set of theoretical commitments about meaning, development, judgement and educational relationships. To understand assessment dialogically is to understand that grades, comments, criteria and standards do not acquire educational meaning as isolated transmissions. They become meaningful through uptake, response, negotiation and use over time.

This section brings together four traditions: Bakhtinian dialogism, Vygotskian sociocultural theory, dialogic pedagogy and social constructivist assessment theory. These traditions differ in intellectual origin and emphasis. Bakhtin provides a theory of utterance, addressivity and the co-creation of meaning. Vygotsky provides a developmental account of mediation, internalisation and the social origins of higher mental functions. Freire, Buber, Alexander, Mercer and Wegerif provide educational accounts of dialogue as co-inquiry, ethical relation and shared thinking. Social constructionist and social constructivist perspectives, including Gergen's relational theory and the Oxford Brookes assessment tradition associated with Rust, O'Donovan and Price, help explain how standards are community-held and learned through participation rather than simply transmitted through criteria. Together, these traditions supply conceptual architecture for dialogic assessment.

1. Bakhtin: Utterance, Addressivity and the Meaning of Feedback

Bakhtin's (1986) dialogism begins with a simple but radical proposition: language is always addressed. An utterance is never merely a neutral container of information. It is spoken or written by someone, to someone, in response to prior utterances and in anticipation of future response. Meaning is therefore not located entirely in the speaker's intention or in the words themselves. It emerges in the relation between utterance, addressee, context and uptake.

This has direct implications for assessment. A teacher's comment "your argument needs more critical engagement" is not a self-sufficient message. It is an utterance addressed to a particular student, situated within a history of teaching, prior feedback, disciplinary expectations and student self-understanding. Its meaning depends on how the student hears it, what prior experiences shape that hearing, what opportunities exist to ask questions, and whether the student can act on it in future work. The comment may be intended as developmental guidance, but it may be received as criticism, as a formula for improvement, as a vague judgement, as confirmation of inadequacy, or as a clue to the assessor's preferences. The educational meaning of feedback therefore lies not in delivery alone but in uptake.

The same is true of grades. A grade appears to be a stable sign: 62, B+, merit, pass, distinction. Yet a grade is also an utterance. It says something about the student's work in relation to a standard, a cohort, a disciplinary expectation and an institutional classification system. It answers the submitted work, but it also anticipates response: acceptance, disappointment, appeal, recalibration, disengagement or renewed effort. A grade that cannot be interpreted educationally functions primarily as a verdict. A grade that becomes part of dialogue can become a resource for understanding standards and future action.

Bakhtin's concept of addressivity is especially important here. Every utterance is shaped by its imagined addressee. Good feedback is not simply accurate; it is addressive. It must be written or spoken with an

understanding of the learner's current interpretive horizon. A technically correct comment may fail educationally if the student cannot understand what action it calls for. Conversely, a brief comment may be powerful if it connects with what the student is ready to see. Feedback quality is thus not only a property of the comment but of the relation between comment and learner.

Bakhtin's (1981) concept of heteroglossia further enriches the analysis. Assessment discourse is multivoiced. A rubric speaks with the voice of the programme, the discipline, the institution and the quality assurance regime. A teacher comment may contain the language of disciplinary expertise, employability, professional standards or institutional policy. A student's response may contain the voice of prior schooling, peer comparison, family expectation, cultural assumptions about authority and anxiety about grades. Assessment is never a simple exchange between two isolated individuals. It is a site where multiple social languages meet, compete and are re-accented.

This explains why assessment misunderstandings are not merely cognitive failures. A student may know the literal meaning of "critical analysis" and still not understand what critical analysis means in a particular discipline. Another may interpret "independent argument" as personal opinion because prior educational experience rewarded assertion. The teacher's words enter a field already populated by other voices. Dialogue is required because meanings do not travel cleanly from assessor to student.

Bakhtin therefore supports a core claim of this book: assessment meaning is co-created in the space between participants. A grade or comment only becomes educationally meaningful in relation to response. Without response, feedback remains monologic. It may be recorded, archived and counted, but it has not necessarily entered learning.

2. Vygotsky: Mediation, ZPD and the Development of Evaluative Judgement

If Bakhtin explains why assessment meaning is relational, Vygotsky explains why assessment development is social. Vygotsky's (1978) sociocultural theory begins with the claim that higher mental functions develop first between people and only later within the individual. The child's capacity to reason, regulate attention, use concepts and solve problems is formed through mediated social activity. What is initially interpsychological becomes intrapsychological.

This principle is crucial for assessment. Evaluative judgement, the capacity to judge the quality of one's own work and that of others, is not born fully formed inside the learner. It develops through participation in evaluative practices. Students first learn what quality means by seeing teachers and peers make judgements, by comparing examples, by hearing reasons, by applying criteria, by receiving response and by revising work. Only gradually can they internalise standards and regulate their own performance.

The zone of proximal development gives this developmental account its pedagogical precision. The ZPD refers to the distance between what a learner can do independently and what the learner can do under adult guidance or in collaboration with more capable peers. Applied to assessment, the relevant question is not only "What can the student currently produce?" but "What kind of judgement can the student begin to exercise with mediation?" This reframes assessment from a record of completed attainment to a site of emerging capability.

Feedback, in this view, is a mediational tool. It does not simply report a gap between current and desired performance. It helps the learner act within that gap. A comment, exemplar, rubric, question, peer discussion or tutorial conversation can mediate development by directing attention, supplying language, modelling criteria and enabling the student to attempt a more advanced form of judgement than would be possible

alone. The value of feedback therefore depends on whether it is usable within the learner's ZPD.

Consider the common practice of giving students exemplars. A measurement view may treat exemplars as illustrations of grade levels. A Vygotskian view treats them as mediational objects around which evaluative dialogue can occur. Students compare two introductions, discuss why one frames the problem more effectively, identify how evidence is used, and attempt to apply those insights to their own drafts. Through this joint activity, standards move from external social discourse toward internalised evaluative capacity.

Peer assessment can be interpreted in the same way. Its value is not only that students receive more comments. Its deeper value is that students practise judgement in a social setting. They must articulate reasons, encounter alternative interpretations, test criteria against actual work and negotiate what counts as quality. In doing so, they begin to appropriate the language and habits of judgement. The developmental case for dialogic assessment is therefore strong: the capabilities assessment claims to measure are often formed through assessment dialogue itself.

3. Dialogic Pedagogy: Freire, Buber, Alexander, Mercer and Wegerif

The dialogic-pedagogy tradition adds a normative and pedagogical dimension. Freire's (1970) critique of the "banking model" of education is directly relevant to assessment. In banking education, teachers deposit knowledge into passive students, and students are later evaluated according to their capacity to reproduce what has been deposited. A parallel banking model of assessment treats feedback as something deposited into students after judgement has occurred. The teacher knows, judges and tells; the student receives, accepts and adjusts.

Freire's alternative is problem-posing education, in which teachers and students become co-investigators of reality. Applied to assessment, this does not abolish teacher expertise or standards. Rather, it changes the

relation through which standards are encountered. The assessed task becomes a problem to be investigated, not a hidden demand to be decoded. Feedback becomes part of co-inquiry: What is this work trying to do? How does it position evidence? What assumptions shape it? What possibilities remain undeveloped? The student is not merely informed of deficiency but invited into investigation.

Buber's (1970) distinction between I–It and I–Thou relations deepens the ethical dimension. Assessment easily becomes I–It: the student is treated as a performance to be measured, a script to be marked, a number to be processed. Dialogic assessment requires an I–Thou orientation, not in the sentimental sense that all judgement must be warm or agreeable, but in the ethical sense that the student is encountered as a developing person capable of response. The work is judged, but the person is not reduced to the judgement. This matters because assessment is one of the most identity-sensitive practices in education. A grade can shape a student's sense of belonging, capability and future possibility. Dialogue protects assessment from becoming purely objectifying.

Robin Alexander's (2008) account of dialogic teaching offers a practical educational framework. Dialogic teaching uses talk to stimulate thinking, extend understanding, build arguments and enable students to participate in shared inquiry. Its relevance to assessment is clear: if learning is advanced through shared reasoning in talk, then assessment should not be confined to the silent exchange of submitted work and returned marks. Assessment events should create opportunities for students to explain reasoning, test interpretations, justify choices and revise understanding.

Mercer's (2000) notion of "interthinking" is equally important. Thinking is not only inside individual heads; it is often accomplished through language used jointly. In group discussion, peer review and teacher-student conferences, students can construct understandings that none might have produced alone. Assessment dialogue can therefore support collective reasoning about quality. Students learn not only from

the teacher's judgement but from the public examination of reasons: Why is this claim stronger? What makes this evidence relevant? Why does this conclusion feel unearned? Such questions make standards visible as reasoning practices.

Wegerif's (2013) work on dialogic space adds the idea that dialogue opens a space between perspectives. Learning occurs when participants encounter difference, hold alternatives in tension and generate new possibilities. Assessment needs such space because quality is rarely obvious. Students must learn to dwell with ambiguity: multiple possible arguments, competing interpretations, uncertain standards, alternative ways to structure a response. A monologic rubric closes uncertainty too quickly. Dialogic assessment teaches students how to reason within it.

4. Social Constructivism and Assessment Standards

Social constructivism in assessment makes explicit what the previous traditions imply: standards are not fixed objects transmitted from teacher to student. They are community-held, historically developed and negotiated in practice. This does not mean that standards are arbitrary or merely subjective. It means that their meaning depends on the interpretive community that sustains them.

Gergen's (2009) relational view challenges individualistic accounts of knowledge and judgement. Meaning is generated in relationship; what counts as knowledge, evidence, competence or quality is produced through social coordination. Applied to assessment, this means that standards are not simply located in the assessor's mind or in the rubric. They exist in the relational practices of a disciplinary community: in seminars, supervision, peer review, marking meetings, exemplars, professional debates and traditions of critique.

The Oxford Brookes social constructivist process model associated with Rust, O'Donovan and Price (2005) is especially significant for assessment theory. It argues that students' understanding of standards develops through active participation in assessment processes. Written

criteria alone are insufficient because much assessment knowledge is tacit. Students need opportunities to engage with exemplars, discuss criteria, apply standards to work, receive feedback and participate in the community's evaluative practices. Standards are thus learned as situated meanings, not as definitions memorised from a handbook.

This model directly challenges the transmission theory of criteria. In the transmission model, teachers define standards, write them into rubrics and deliver them to students. In the social constructivist model, criteria are starting points for shared meaning-making. A criterion such as "critical engagement with literature" becomes meaningful when students examine actual texts, compare stronger and weaker uses of sources, hear disciplinary reasons and attempt application in their own writing. The standard is not hidden, but neither is it fully explicit. It becomes knowable through guided participation.

The implication for the book is decisive: dialogic assessment is not an optional enhancement to assessment. It is the process through which standards become educationally available. Without opportunities to interpret and apply them, standards risk remaining private to assessors or appearing falsely self-explanatory in documents. Dialogue is the means by which community-held judgement becomes shareable.

5. A Working Definition of Dialogic Assessment

For the purposes of this book, dialogic assessment is defined as a family of practices in which judgement, standards and feedback are co-constructed through reciprocal, response-able exchange between students and teachers, and among students, over time.

Several elements of this definition require emphasis. First, dialogic assessment is a family of practices, not a single technique. It may include feedback conferences, peer review, self-assessment, exemplar analysis, draft dialogue, negotiated criteria, oral defence, portfolio conversation, dialogic marking, reflective commentary and iterative resubmission. What

unites these practices is not their format but their orientation toward reciprocal meaning-making.

Second, judgement is co-constructed. This does not mean that students and teachers always have equal authority over grades. Summative responsibility may remain with the assessor. Co-construction means that the meaning of judgement is developed through exchange: reasons are offered, interpretations are tested, standards are discussed and future action is clarified. The grade may be awarded by the teacher, but its educational meaning is not complete until it is understood and used by the learner.

Third, standards are co-constructed. This does not imply that standards are invented anew in every conversation. Rather, disciplinary standards become intelligible through situated dialogue. Students are gradually inducted into what the community values and how it reasons about quality.

Fourth, feedback is response-able. The hyphen matters. Dialogic feedback is not merely feedback to which a student could theoretically respond. It is designed so that response is expected, enabled and consequential. Students must have opportunities to question, apply, revise, contest and carry feedback forward. Otherwise, feedback remains delivery.

Finally, dialogic assessment unfolds over time. A single conversation may be useful, but dialogue as an assessment principle requires temporal structure: before submission, during production, after judgement and into subsequent work. Standards are not internalised in one event. Evaluative judgement develops through repeated participation.

6. What Dialogue Is Not

Because “dialogue” is often used loosely, it is necessary to distinguish dialogic assessment from adjacent but weaker practices. Dialogue is not mere conversation. A friendly exchange between teacher and student may improve rapport without altering the structure of assessment meaning. Dialogue requires orientation toward shared inquiry, reasons, standards and response.

Dialogue is not consultation. Asking students for preferences about deadlines or topics may be valuable, but consultation can leave the deeper assessment structure untouched. Students may be heard without becoming participants in judgement. Dialogic assessment requires participation in the meaning of quality, not only in procedural decisions.

Dialogue is not feedback delivery. Even detailed feedback remains monologic if the student has no opportunity to interpret it with others, act on it, or test its meaning in subsequent work. A comment bank, however efficient, is not dialogic by itself. Nor is audio feedback automatically dialogic. Modality does not determine dialogue; reciprocity does.

Dialogue is also not the abandonment of expertise. A dialogic assessment system does not pretend that students already understand standards as experts do. It recognises asymmetry but refuses to make asymmetry absolute. Teachers bring disciplinary authority; students bring emerging understanding, questions, intentions and interpretations. Dialogue is the structured encounter between these positions.

Finally, dialogue is not vagueness. On the contrary, dialogic assessment requires precision: precision about criteria, purposes, roles, evidence, reasons and next steps. Its critique of measurement is not a rejection of clarity but a rejection of the belief that clarity can be achieved by documents alone.

1.4 Scope, Aims, and Guiding Questions

This book is organized around a single overarching question: What would it mean to design and enact assessment in higher education as dialogue rather than as one-way measurement, and what are the implications for students, educators, institutions, and the idea of academic standards? The question is deliberately theoretical and practical at once. It asks not only how assessment might be differently conceptualized, but also how reconceptualization might alter the design of tasks, the use of feedback, the interpretation of standards, the responsibilities of teachers and students, and the institutional conditions under which assessment is made accountable.

The book starts from the premise that contemporary higher education assessment is marked by a persistent contradiction. On the one hand, the literature on assessment for learning, feedback literacy, evaluative judgement and student partnership has substantially challenged the idea that assessment is only a mechanism for measuring attainment. On the other hand, institutional assessment systems continue to be organized largely around summative verification, documentary accountability, grades, rubrics, and defensible procedures. The problem, therefore, is not a lack of reform vocabulary. Higher education already has an extensive language of formative assessment, dialogic feedback, sustainable assessment and learner agency. The problem is that these ideas often remain peripheral to the core architecture of assessment systems. This book seeks to bring them into a more coherent theoretical and practical framework by placing dialogue at the centre.

1. Research Questions

The overarching question is developed through five subsidiary questions, each of which foreshadows a chapter of the book.

First, how did assessment-as-measurement become the dominant paradigm of higher education assessment, and what are the limits of relying on measurement alone? This question frames Chapter 1. It

requires an examination of the historical and conceptual authority of measurement, the rise of criteria-based assessment, and the appeal of rubrics as technologies of transparency and control. It also requires attention to the communicative failures of contemporary assessment: weak feedback uptake, strategic compliance, grade anxiety, and the erosion of trust.

Second, what would an assessment architecture look like if dialogue were treated not as an optional supplement but as a design principle? This question frames Chapter 2. It concerns the structural design of assessment systems, including timing, sequencing, task design, feedback loops, opportunities for revision, peer interaction, self-assessment and the development of evaluative judgement. The aim is to move from isolated dialogic practices toward a programme-level architecture.

Third, how can selected research inform dialogic assessment designs that support learning, engagement and judgement? This question frames Chapter 3. The chapter uses illustrative designs informed by published research to examine possible benefits, conditions and limitations. These examples do not constitute original case-study evidence or establish universal causal effects.

Fourth, what theoretical account of standards, judgement and feedback is required to sustain assessment as dialogue? This question frames Chapter 4. It develops the conceptual framework more fully by drawing together dialogism, sociocultural theory, social constructivism, feedback literacy and evaluative judgement. The chapter asks how academic standards can be understood as socially held and dialogically learned without collapsing into relativism.

Fifth, what are the implications of dialogic assessment for students, educators, institutions and the future governance of academic standards? This question frames Chapter 5. It considers workload, equity, staff development, student responsibility, quality assurance, digital assessment environments and institutional policy. The chapter asks how dialogue can

be made sustainable within real higher education systems rather than remaining an ideal attached to small-scale innovation.

2. Scope and Boundaries

The book focuses on higher education, including undergraduate and postgraduate taught provision, with some relevance to research supervision and professional education. Its examples are drawn across disciplines because the central problem is not discipline-specific. Essays, laboratory reports, studio critiques, clinical performances, dissertations, portfolios, reflective writing and examinations all raise questions about how standards are communicated, how judgement is formed and how students learn to interpret feedback. Nevertheless, the book recognizes that assessment practices are always discipline-sensitive. Dialogue in studio-based assessment does not look identical to dialogue in engineering design, literary interpretation, nursing practice, sociology essays or doctoral supervision.

The book combines conceptual argument with selected research and illustrative designs. It is not an empirical report of a single institutional intervention. Rather, it is a selective conceptual synthesis of theoretical traditions, research literature and illustrative cases, not a systematic review. It proposes an integrative framework. The illustrative designs used throughout the book are analytical rather than exhaustive. They are intended to show how dialogic principles might operate in concrete assessment settings, not to claim universal causal effects.

The literature base is predominantly Anglophone. The selected literature draws mainly on English-language research from the United Kingdom, Australia, New Zealand, Canada, the United States and parts of Europe. This is a limitation. Assessment cultures vary significantly across national systems, institutional histories and linguistic contexts. The book does not claim that Anglophone debates exhaust the global field of assessment. Rather, it uses them as the principal site through which the measurement-dialogue tension is examined, while acknowledging that

future work should extend the framework through comparative and multilingual scholarship.

The book also has clear exclusions. It is not a psychometrics manual. It does not offer technical instruction in test construction, item response theory, standard setting, statistical reliability or measurement modelling. These fields are important, but they are not the object of this study. Nor is the book a step-by-step guide to rubric writing. It discusses rubrics critically because they are central to contemporary assessment practice, but its purpose is not to provide templates for analytic or holistic rubrics. Similarly, the book is not a general handbook on feedback techniques. Its concern is more fundamental: the conceptual and structural conditions under which feedback, standards and judgement become dialogic.

3. Structure of the Book

Chapter 1 develops the critique. It argues that higher education assessment remains governed by a measurement paradigm that privileges summative certification over formative development and self-regulatory learning. It examines the rubric as the emblematic technology of this paradigm and establishes dialogue as the book's theoretical alternative. The chapter concludes with the present scope-and-aims statement.

Chapter 2 develops the architecture. It asks what assessment design would require if dialogue were built into the structure of courses and programmes rather than added after marking. It examines sequencing, feedforward, iterative tasks, peer assessment, self-assessment, exemplars, dialogic rubrics, feedback conferences and opportunities for revision. The chapter's central contribution is to translate the theoretical claim into design principles.

Chapter 3 turns to practice. It develops illustrative designs informed by research on dialogic feedback, student self-regulation, peer review, assessment literacy, evaluative judgement and partnership approaches. It considers their possible benefits and limitations without presenting the examples as original empirical studies.

Chapter 4 develops the theory of standards and judgement. It argues that academic standards are neither wholly fixed objects nor merely subjective preferences. They are socially held, historically developed and learned through participation in communities of practice. This chapter brings together sociocultural theory, dialogism, social constructivism and judgement theory to explain how standards can be both rigorous and dialogically shared. It culminates in the proposed Dialogic Assessment Framework, a conceptual synthesis requiring empirical evaluation.

Chapter 5 considers implications. It asks what a dialogue-centred assessment framework would mean for students, educators, institutions and quality assurance. It addresses practical constraints including workload, scalability, class size, digital platforms, appeals, policy, staff development and equity. The final chapter argues that the future of assessment reform depends not on abandoning standards, but on redesigning the conditions through which standards are encountered, interpreted and enacted.

The argument therefore builds in five movements: from critique, to architecture, to illustrative practice, to theory, to implications. The book begins by diagnosing the limitations of assessment-as-measurement, then develops a dialogic alternative, explores its application through research-informed examples, strengthens its theoretical foundations and finally considers its institutional consequences.

4. Intended Contribution

The book contributes to higher education assessment scholarship in three ways. First, it integrates bodies of literature that are often discussed separately: assessment for learning, feedback literacy, evaluative judgement, dialogic pedagogy, social constructivism and standards-based assessment. Each of these fields has already challenged transmissive models of assessment, but they have not always been brought together under a single conceptual framework. This book proposes dialogue as that integrative centre.

Second, it extends the feedback-literacy and evaluative-judgement literature by arguing that students' capacity to use feedback and judge quality is not merely an individual skill. It is a relational and developmental achievement formed through structured participation in assessment practices. Students become capable of judgement not simply by receiving better information, but by entering conversations in which standards are modelled, interpreted, contested and applied.

Third, the book offers a translational agenda for practice. It does not stop at critique, nor does it present dialogue as an abstract ideal. It asks how dialogic principles can be designed into assessment ecologies: through task sequencing, exemplars, peer dialogue, feedback uptake, revision cycles, teacher judgement, programme design and institutional policy. The intended contribution is therefore both theoretical and practical: an account of assessment as a communicative, relational and standards-oriented practice, and a framework for redesigning higher education assessment accordingly.

The central claim is that dialogue is not opposed to rigour. Properly understood, it is a condition of rigour. Standards cannot remain rigorous if students experience them only as opaque rules, grades or rubrics. They become educationally rigorous when students are inducted into the practices of judgement through which standards have meaning. This book therefore seeks to move beyond the rubric not by abandoning explicitness, fairness or accountability, but by re-situating them within the dialogic life of higher education.

The Architecture of Dialogic Assessment

Principles, Designs, and Enabling Conditions

2.1 Defining Dialogic Assessment: Core Principles and Boundary Conditions

The argument of this book depends on a precise definition of dialogic assessment. The term cannot be allowed to function as a vague virtue-word, a general synonym for good teaching, or a decorative label attached to any practice in which students are invited to speak. Dialogue is attractive partly because it appears ethically generous: it suggests participation, openness, reciprocity and respect. Yet precisely because of this attractiveness, it is vulnerable to dilution. If every feedback comment, class discussion, consultation survey or opportunity for students to ask questions is described as dialogic, then the concept loses analytical force. A rigorous account must specify what makes assessment dialogic, what distinguishes it from adjacent assessment traditions, where its limits lie, and how it operates under the persistent asymmetry between assessor and assessed.

Dialogic assessment, as used in this book, refers to a family of assessment practices in which judgement, standards and feedback are constructed, interpreted and revised through reciprocal, response-able exchange between students and teachers, and among students, over time. It is not a single method. It may include feedback conferences, dialogic

feedforward, peer review, self-assessment, exemplar discussion, oral defence, draft dialogue, negotiated criteria, portfolio review or collaborative moderation of standards. What unites these practices is not their format but their underlying logic: assessment is treated not as the one-way transmission of judgement from assessor to student, but as a socially situated process through which students learn to understand, use and eventually exercise judgement.

1. Core Principles of Dialogic Assessment

The first principle is reciprocity. In dialogic assessment, both parties contribute to the making of meaning, and both may be changed by the exchange. This does not imply equality of expertise or authority. The teacher may still possess greater disciplinary knowledge and may retain formal responsibility for summative judgement. Reciprocity means that the student is not positioned merely as the receiver of judgement. The student brings intentions, interpretations, uncertainties, questions, self-evaluations and responses that matter to the assessment process. A feedback conversation in which the teacher explains the mark while the student silently listens is not fully dialogic. A conversation in which the student explains what they were trying to do, identifies where they were uncertain, tests the teacher's interpretation, and uses the exchange to reframe future work moves closer to dialogic assessment.

The second principle is response-ability. The hyphen is important. Dialogic assessment requires that utterances are made to be taken up. A grade, comment, question or criterion is not complete at the moment of delivery. It is educationally complete only when it enters student action, interpretation or further dialogue. Feedback must therefore be designed with response in view. Students require opportunities to ask what a comment means, to apply it to a revision, to compare it with other evidence, to challenge it where appropriate, and to carry it into subsequent tasks. A system that gives extensive comments after the final submission but offers no occasion for uptake remains largely monologic, however detailed the feedback may be.

The third principle is temporality. Dialogic assessment unfolds over time. It cannot be reduced to the endpoint at which a performance is marked. Judgement develops through cycles: anticipation, production, response, interpretation, revision and transfer. Students encounter standards before submission through task discussion, exemplars and criteria; during production through questions, drafts and peer review; after submission through feedback, grades and reflection; and beyond the task through feedforward into future work. The temporal structure matters because evaluative judgement is not acquired instantly. Students need repeated participation in assessment-related dialogue in order to internalize standards and develop independence.

The fourth principle is the co-construction of standards. This does not mean that standards are invented from nothing by each cohort or negotiated as matters of personal preference. Higher education standards are embedded in disciplines, professions and institutional histories. Yet students do not gain access to those standards simply by reading rubrics or learning outcomes. Standards become meaningful through use: comparing examples, hearing reasons, applying criteria, noticing differences in quality, and discussing borderline cases. Co-construction means that teachers and students work together to make standards intelligible in context. The standard is not surrendered; it is shared through guided participation.

The fifth principle is the cultivation of student evaluative judgement. Dialogic assessment is not primarily about improving the emotional tone of feedback, although that may be one consequence. Its aim is development: students should become increasingly able to judge the quality of their own work and the work of others. This capability is central to higher education because graduates must continue to make judgements beyond formal instruction. They must evaluate evidence, decide whether work is good enough, seek feedback, interpret competing standards and improve performance in unfamiliar settings. Dialogic

assessment is therefore justified not only because it makes feedback more humane, but because it develops the capacity for independent judgement.

The sixth principle is attention to relational and affective conditions. Assessment is never emotionally neutral. Grades can provoke pride, shame, resentment, relief, defensiveness or disengagement. Feedback may be heard as guidance, accusation, encouragement or rejection depending on the relationship in which it is received. Dialogic assessment requires trust, but not uncritical comfort. It requires conditions in which students can ask questions without appearing deficient, disclose uncertainty without fear of punishment, and encounter critique as part of learning rather than as a judgement on their worth. The affective dimension is not a soft addition to assessment; it shapes whether feedback can be heard, interpreted and used.

2. Distinguishing Dialogic Assessment from Adjacent Concepts

Dialogic assessment overlaps with several influential traditions in assessment scholarship, but it is not identical with any of them. The first is formative assessment. Formative assessment refers to assessment used to inform and improve learning while learning is still in progress. It is defined by function: evidence is gathered and used to modify teaching or learning. Dialogic assessments can be formative, but not all formative assessments are dialogic. A teacher may identify a misconception through a quiz and adjust the next lecture without any substantial student participation in judgement. This is formative, but not necessarily dialogic. Dialogic assessment requires reciprocal meaning-making, not only instructional adjustment.

The second adjacent concept is feedback. Feedback is often defined as information about performance that helps close the gap between current and desired achievement. Dialogic assessment includes feedback, but it rejects the idea that feedback is simply information delivered by a knowledgeable source. In dialogic assessment, feedback is a process of sense-making and action. A written comment is only one moment in that

process. The key question is not “Was feedback provided?” but “What did the student understand, negotiate, question, use and carry forward?” Feedback can remain monologue if it is delivered without uptake.

The third concept is Carless’s (2007) learning-oriented assessment. Learning-oriented assessment emphasizes assessment tasks as learning tasks, student involvement in assessment, and feedback processes that support future learning. Dialogic assessment shares this orientation strongly. It also agrees that assessment design should generate learning rather than merely measure it. The distinction is that dialogic assessment places particular analytical emphasis on the communicative and relational processes through which judgement and standards become meaningful. Learning-oriented assessment is a broader design framework; dialogic assessment specifies the exchange conditions under which that design becomes educative.

The fourth concept is Boud’s (2000) sustainable assessment. Sustainable assessment asks whether assessment develops learners’ capacity to judge and regulate their own learning beyond the immediate course. Dialogic assessment shares this future-facing aim. Its cultivation of evaluative judgement is explicitly sustainable in this sense. The distinction is again one of emphasis. Sustainable assessment foregrounds the long-term capacity of learners to make independent judgements. Dialogic assessment foregrounds the social and interactional processes through which such independence is developed. It argues that independence is not produced by leaving students alone with standards, but by engaging them in structured dialogue until judgement can be internalized.

The fifth concept is assessment-as-learning. Assessment-as-learning emphasizes students’ metacognitive and self-regulatory engagement with assessment. Students monitor their own learning, use criteria and reflect on progress. Dialogic assessment includes this aim but resists an overly individualistic interpretation. Self-regulation is not simply a private mental skill. It is formed through social interaction. Students learn to regulate judgement by first participating in shared judgement. Thus,

dialogic assessment treats assessment as a social achievement before it becomes an individual capacity.

3. Boundary Conditions and Limits

Dialogic assessment is not appropriate in the same way, to the same degree, or at every moment of assessment. Its limits must be acknowledged if the concept is to remain credible. The first limit concerns high-stakes certification. Some assessment decisions require formal independence, procedural consistency and clear authority. Final degree classifications, professional licensing examinations, clinical competency decisions and safety-critical judgements cannot be fully negotiated with students. A nursing student's competence in medication administration, an engineer's capacity to meet safety standards, or a teacher's fitness for professional practice cannot be decided through open-ended dialogue alone. In such cases, dialogue may support preparation, interpretation and feedback, but the final judgement must retain gatekeeping force.

The second limit concerns professional gatekeeping. Higher education does not only support learning; it also protects public trust. In fields such as medicine, law, education, engineering and social work, assessment carries responsibilities to clients, patients, pupils and communities. Dialogic assessment must not be confused with lowering standards or allowing students to bargain their way around competence requirements. Its task is to make standards more intelligible and developmental, not less demanding.

The third limit concerns scale. Large cohorts impose constraints on sustained individual dialogue. A module with five hundred students cannot easily provide repeated one-to-one feedback conferences. However, scale does not make dialogue impossible. It requires different designs: peer dialogue, group feedback, annotated exemplars, structured self-assessment, digital forums, short feedback-response tasks, calibration exercises and teaching assistants trained in shared standards. The question is not whether every assessment can become a seminar-like conversation,

but whether the assessment ecology creates structured opportunities for meaningful uptake and participation.

The fourth limit is the risk of dialogue-washing. This occurs when institutions adopt the language of dialogue, partnership or student voice while leaving power relations and assessment structures untouched. A module may ask students whether they understood the rubric but give them no role in interpreting standards. A programme may consult students on assessment design after the major decisions have already been made. A teacher may invite questions but treat disagreement as evidence of student misunderstanding. Such practices create the appearance of participation without redistributing interpretive agency. Dialogue-washing is especially likely where institutions seek the legitimacy of student-centred language without altering workload models, marking practices, appeals processes or curriculum structures.

The fifth limit is that dialogue requires capacity. Students may not initially know how to participate in assessment dialogue. They may ask only “What do I need to do to get a higher mark?” or seek certainty where judgement requires ambiguity. Teachers may also lack dialogic assessment expertise. They may dominate conversation, over-explain, become defensive, or interpret student questioning as challenge rather than engagement. Dialogic assessment therefore requires induction, norms, modelling and professional development. It is not achieved merely by creating opportunities to talk.

4. A Spectrum of Assessment Communication

Dialogic assessment can be understood as part of a spectrum. At one end is monologic assessment. Here judgement moves one way: from assessor to student. The student submits work; the teacher marks it; the grade and comments are returned; the process ends. A typical example is the final examination in a modular course. Students receive a mark weeks later, perhaps with generic cohort feedback, but no meaningful opportunity to discuss, revise or carry forward the judgement. The assessment may be valid for certification, but its communicative structure is closed.

A second position on the spectrum is responsive assessment. Here feedback invites reply, but judgement itself remains largely assessor controlled. For example, a teacher may return essays with comments and allow students to submit a brief response explaining how they will use the feedback in the next assignment. Or a lecturer may hold a post-assignment workshop in which students ask questions about common strengths and weaknesses. These practices improve uptake because they create channels for student response. However, the standards and judgement remain mostly transmitted from teacher to student. Responsive assessment is stronger than monologue but not yet fully dialogic.

A third position is dialogically supported assessment. Here students participate in the interpretation of standards before or during the task. For example, in a postgraduate research methods module, students might examine three sample literature reviews, compare them against criteria, discuss disagreements in groups, and then produce a draft section for peer and tutor discussion. The final mark may still be awarded by the tutor, but the student has participated in the construction of evaluative meaning before submission. This design supports judgement development rather than merely feedback reception.

At the far end is fully dialogic assessment, where judgement is co-constructed within structured limits. Consider a capstone portfolio in which students select evidence, write a self-evaluative commentary, discuss their claims with peers, receive tutor questions, revise the

portfolio, and participate in an oral defence. The assessor retains final responsibility for the mark, but the judgement emerges through a process in which students must articulate standards, justify evidence, respond to critique and demonstrate evaluative understanding. Here assessment is not simply applied to work; it becomes part of the work of learning.

This spectrum is analytically useful because it avoids a binary division between dialogic and non-dialogic assessment. Many practices are hybrid. A final examination may be monologic in judgement but dialogically prepared through prior workshops. A rubric may be monologic if simply distributed, responsive if students can ask questions, and dialogic if used with exemplars, peer calibration and revision. The issue is not the presence of a tool but the communicative relation it enables.

5. Dialogue and Power

The most serious objection to dialogic assessment is that assessment is inherently unequal. Teachers grade; students are graded. Institutions award credentials; students seek them. Assessors control standards, deadlines, criteria and consequences. Dialogue cannot abolish this asymmetry, and it should not pretend to do so. A teacher who says “we are all equal here” while retaining the authority to award marks may conceal power rather than transform it.

The claim of dialogic assessment is more modest and more defensible: dialogue can make asymmetry visible, accountable and educative. It makes power visible by acknowledging who has authority to judge and why. It makes power accountable by requiring reasons, opportunities for clarification, and responsiveness to student interpretation. It makes power educative by using the asymmetry between expert and novice as a resource for development rather than as a mechanism of domination.

This distinction is crucial. Expertise is not the enemy of dialogue. Students often need access to the teacher’s more developed judgement. The problem arises when expertise is exercised as unanswerable authority. Dialogic assessment asks teachers to make their judgement available for

learning: to explain reasons, model standards, invite student attempts at judgement, and respond to how students understand the assessment. The teacher remains an assessor but also becomes a mediator of judgement.

For students, dialogue does not mean that every preference must be accepted. It means that their interpretations matter educationally. A student who misunderstands a standard should not simply be penalized for misunderstanding; the misunderstanding should become a site of learning. A student who disagrees with feedback should not automatically prevail, but the disagreement may reveal ambiguity in the task, hidden assumptions in the criteria, or a need for clearer modelling. Dialogue therefore changes the meaning of student voice. It is not consumer feedback. It is participation in the work of judgement.

The power question also guards against romanticizing dialogue. Some students may feel less able to speak because of language background, prior educational experience, disability, race, class, gender, age, cultural norms or fear of appearing incompetent. Dialogic assessment must therefore be designed inclusively. It cannot rely only on confident students speaking in open discussion. It requires multiple modes of participation: written questions, anonymous polling, structured peer roles, small-group dialogue, reflective commentaries, digital forums and opportunities for private clarification.

2.2 Feedback as Dialogue, Not Delivery: From Information to Interaction

Few practices in higher education have generated as much reform rhetoric as feedback. Students ask for more of it; teachers report spending substantial time producing it; institutions monitor it through student satisfaction surveys; quality assurance documents present it as central to learning. Yet the effectiveness of feedback is highly variable. The problem is not simply that feedback is absent, late or insufficiently detailed, although each of these problems is real. The deeper problem is conceptual. For much of its institutional life, feedback has been imagined as information given by an expert to a novice after performance. The teacher comments, the student receives, and learning is expected to follow. This is the transmission or delivery paradigm of feedback.

The central shift in contemporary feedback theory is away from this delivery model and toward an uptake-centred, dialogic model. In the new paradigm, feedback is not defined by what teachers provide but by what learners make of information and do with it. A comment sitting unread in a learning management system is not yet feedback in any meaningful educational sense. A rubric annotation that students cannot interpret is not feedback. A grade accompanied by extensive justification is not feedback if it produces no understanding, judgement or future action. Feedback exists as a process of sense-making, comparison, response and use. It is therefore inseparable from dialogue, even when the dialogue is distributed across texts, peers, exemplars, self-assessment, digital tools and subsequent tasks rather than occurring only in face-to-face conversation.

1. The Old Paradigm: Feedback as Expert Transmission

The old paradigm conceives feedback as corrective information transmitted from teacher to student. Its implicit model is linear. The student produces work; the teacher evaluates it; the teacher identifies strengths and weaknesses; the teacher writes comments; the student is expected to improve. The teacher is positioned as the source of judgement, the student as the recipient of judgement, and the written comment as the vehicle through which improvement is supposed to occur.

This model is attractive because it corresponds to common institutional routines. It is administratively tidy. It locates responsibility for feedback in the assessor's production of comments. It can be audited: were comments provided, how many, how quickly, in what format? It also fits the summative structure of modular higher education. Students submit work near the end of a unit; teachers mark it; comments are returned; the module closes; both parties move on. Within this structure, feedback becomes a postscript to judgement rather than a condition of learning.

Nicol and Macfarlane-Dick's (2006) intervention was important because it challenged the assumption that feedback is mainly something teachers give. By placing feedback within a model of self-regulated learning, they argued that students are not empty receivers of comments but active agents who must monitor, interpret and regulate their own learning. Feedback should help students understand goals, compare current performance with desired performance and take action to close the gap. This already destabilizes the delivery model: if feedback supports self-regulation, then its value cannot be assessed by the existence of comments alone.

Carless's (2006) work on differing perceptions in the feedback process similarly exposed the instability of the transmission model. Teachers and students often understand feedback differently. Teachers may believe that their comments are developmental, while students may see them as

justification for a mark. Teachers may assume that students can interpret disciplinary shorthand; students may experience comments as vague, contradictory or emotionally discouraging. Teachers may complain that students are grade-oriented, while students may respond that the assessment system itself teaches them to be grade-oriented. The problem is not simply poor communication but a mismatch in the meanings attached to feedback.

Boud and Molloy (2013) distinguish information delivery from “feedback Mark 1,” a teacher-driven process in which information about current work is used to improve subsequent performance and the effect is checked. Mark 1 therefore already requires a feedback loop; it is not simply the transmission of comments. Its limitation is its dependence on the teacher to initiate and monitor that loop. This distinction matters because extensive commenting alone may leave student uptake under-designed.

Research identifies recurrent barriers to feedback uptake (Carless, 2006; Carless & Boud, 2018). Students frequently attend first to the grade, with comments becoming secondary or irrelevant once the grade has determined emotional response. Some students read comments only when disappointed by the mark; others do not collect feedback at all if no further task requires its use. When comments are read, they may not be understood. Phrases such as “more critical analysis,” “develop your argument,” “engage more deeply with the literature,” or “improve coherence” may be meaningful to assessors but opaque to students who have not yet internalized disciplinary standards. Even when students understand the general criticism, they may not know what action would count as improvement.

The old paradigm therefore fails not because teachers are careless, but because it misdescribes the learning process. It assumes that the main challenge is producing better comments. This leads to familiar technical solutions: more detailed annotations, standardized feedback sheets, faster turnaround, comment banks, audio comments, marking codes and digital

rubrics. These may improve the conditions of feedback, but they do not by themselves solve the problem of uptake. A clearer comment remains educationally inert if students cannot interpret it, connect it to standards, manage its emotional impact or apply it to future work.

2. The New Paradigm: Feedback as Sense-Making and Uptake

The uptake-centred account developed by Carless and Boud (2018) defines feedback as a process in which learners make sense of information about their performance and use it to enhance the quality of their work or learning strategies. This definition shifts the centre of gravity from teacher provision to learner action. Feedback is not the same as comments. Comments may be one source of feedback information, but they become feedback only when students use them in making judgements and decisions.

Boud and Molloy's (2013) "feedback Mark 2" captures this shift. Feedback is no longer a post-performance message delivered by the teacher, but a designed process that positions learners as active participants. It must be prospective as well as retrospective. It must help students develop capacities for self-monitoring, evaluative judgement and future learning. The design problem is therefore not "How can teachers give more feedback?" but "How can assessment create conditions in which students generate, interpret and use feedback?"

This has several implications. First, feedback must be connected to future action. If students receive comments after the final task in a module with no opportunity to use them, the system has not designed for feedback; it has designed for explanation. Explanatory comments may satisfy accountability, but they do not necessarily support learning. Uptake requires a subsequent task, revision opportunity, reflective response, resubmission, portfolio commentary, or structured transfer into a later assignment.

Second, feedback must involve sense-making. Students need opportunities to interpret comments, compare them with criteria, discuss

them with peers, ask questions and translate them into action. Sense-making is especially important because feedback language is often compressed. A teacher's short phrase may carry years of disciplinary experience. Without dialogue, students must decode that experience alone.

Third, feedback must be distributed. Useful information about performance may come from teachers, peers, exemplars, rubrics, professional standards, prior work, automated systems, audience response or self-comparison. A dialogic model does not restrict feedback to teacher comments. It asks how students are helped to compare their work against meaningful reference points and generate internal feedback from those comparisons.

Fourth, feedback must be understood as relational. Students' willingness to engage with critique depends on trust, belonging, confidence and perceptions of fairness. A harsh but accurate comment may be unusable if it produces defensiveness or shame. A supportive comment may be insufficient if it avoids the difficult judgement required for improvement. Dialogic feedback requires a relational frame in which critique can be demanding without being destructive.

3. Feedback Literacy on Both Sides

The uptake-centred paradigm makes feedback literacy essential. Student feedback literacy, as formulated by Carless and Boud (2018), includes four interrelated capacities: appreciating feedback, making judgements, managing affect and taking action.

Appreciating feedback means understanding its purpose and value. Students must see feedback not as correction imposed by authority but as information that can improve work and develop judgement. This is not automatic. In grade-dominated assessment systems, students may see feedback as an explanation of the mark rather than a resource for learning. Appreciation therefore requires induction into feedback as a process.

Making judgements refers to students' capacity to evaluate the quality of work, including their own. This is central because feedback cannot be used without judgement. Students must decide which comments matter most, how they relate to criteria, whether they apply to future tasks, and what revisions would improve quality. Peer review, exemplar analysis and self-assessment are especially important here because they allow students to practise judgement rather than merely receive it.

Managing affect concerns the emotional dimension of feedback. Students may feel threatened, disappointed, confused or defensive when encountering critique. Feedback literacy includes the capacity to remain engaged despite discomfort. This does not mean making students responsible for absorbing poorly framed feedback. It means designing feedback cultures in which challenge is normalized, support is available, and students learn to interpret critique as part of development rather than as a judgement on identity.

Acting is the decisive feature. Feedback literacy is incomplete if students appreciate feedback, understand it and manage emotion but do not act. Action may involve revising work, changing study strategies, seeking clarification, comparing examples, planning differently, or monitoring a recurring weakness. In an uptake-centred model, action is the point at which feedback becomes real.

However, student feedback literacy cannot be developed without teacher feedback literacy. Carless and Winstone's (2023) framework is useful because it refuses to place responsibility solely on students. Teachers need design literacy, relational literacy and pragmatic literacy.

Design literacy concerns the capacity to construct feedback processes for uptake. Teachers must sequence tasks so that feedback can be used, create opportunities for comparison, build in response points and align feedback with future learning. A teacher who writes excellent comments after a terminal assignment may be feedback-literate at the level of expression but not at the level of design.

Relational literacy concerns how feedback is framed, communicated and emotionally held. Teachers must understand that comments are received by students with histories, anxieties and unequal confidence in academic language. Relational literacy does not mean avoiding rigorous critique. It means making critique usable through trust, respect and clarity of purpose.

Pragmatic literacy concerns the compromises required in real institutions. Teachers work with large cohorts, limited marking time, digital systems, moderation requirements and disciplinary constraints. Feedback literacy therefore includes knowing when not to write more comments, when to use group feedback, when to design peer processes, when to use exemplars, and how to create scalable feedback architectures.

4. Dialogic Feedback Designs

If feedback is defined by uptake, then design must build loops rather than endpoints. A feedback loop requires at least four elements: performance, information, interpretation and subsequent action. Many higher education practices contain the first two elements but omit the last two. Students perform; teachers provide information; the loop remains open. Dialogic design closes the loop by requiring students to interpret and use feedback.

Feed-forward is one common strategy. Instead of treating feedback as commentary on past work, feed-forward connects judgement to a future task. For example, students may receive feedback on a proposal that must be used in the final project. Or feedback from a first essay may be explicitly linked to criteria in a second essay. The point is not merely that comments refer to the future, but that the curriculum creates a future in which the comments matter.

Two-stage assignments provide stronger architecture. Students submit a draft, receive peer and teacher responses, revise, and submit a final version with a commentary explaining how feedback was used. This design changes the student's role. The student is no longer only judged;

the student must judge the feedback, decide what to act on, and justify revisions. The teacher can then assess not only the product but the student's developing evaluative judgement.

Coversheets and self-assessment requests offer a lighter but powerful mechanism. Before submission, students may be asked to identify the aspect of the work on which they most want feedback, state how they have used previous feedback, or evaluate their work against selected criteria. This invites students into the feedback process before the teacher comments. It also helps teachers target feedback where it is most likely to be taken up. A generic comment becomes a response to a student's self-identified concern.

Audio and screencast feedback can support dialogue by making tone, emphasis and reasoning more accessible. A teacher talking through a student's introduction on screen can show how a reader experiences the argument in real time. This may make tacit judgement more visible than marginal comments alone. However, audio or screencast feedback is not automatically dialogic. If students cannot respond, ask questions or use the feedback in subsequent work, the medium changes but the delivery model remains intact.

Feedback tutorials provide a more explicitly dialogic format. These may occur after draft submission, after marking, or at key transition points in a project. Their value lies in allowing the student to explain intentions, ask questions and test interpretations. The tutorial can reveal that the student misunderstood the task, misread the criterion, or had a strong idea but lacked the disciplinary resources to develop it. Such discoveries are unlikely to emerge from written comments alone.

Yang and Carless's (2013) feedback triangle helps organize these designs. The cognitive dimension concerns the content of feedback: the quality of information about standards, performance and improvement. The social-affective dimension concerns trust, emotion, identity and interpersonal negotiation. The structural dimension concerns timing, sequencing, workload, institutional systems and opportunities for action.

Dialogic feedback requires all three. Cognitive precision without relational support may be unusable. Relational warmth without cognitive challenge may be unproductive. Both will fail if structural design gives students no opportunity to act.

5. Internal Feedback and Productive Comparison

Nicol's (2021) work on internal feedback deepens the dialogic paradigm by showing that students generate feedback through comparison. When students compare their work with an exemplar, a peer text, a rubric, a teacher comment, a previous attempt or a professional standard, they produce internal feedback: they notice gaps, possibilities, strengths and alternatives. This reframes feedback as something students generate, not merely something teachers give.

This view is important because it expands the meaning of dialogue. Dialogue is not only interpersonal conversation. It can also be the orchestration of productive comparisons. A student placing their draft beside two exemplars enters a kind of dialogue with those texts. A peer review activity creates dialogue between alternative performances. A reflective coversheet creates dialogue between the student's intention and the assessor's response. A portfolio creates dialogue across time between earlier and later work. Such comparison supports dialogic learning, although it does not by itself provide the reciprocal exchange or shared authority of interpersonal dialogue.

The implication is that teachers should plan comparison opportunities deliberately. Instead of simply telling students what is wrong, teachers can ask students to compare two introductions and identify which better establishes a problem. Instead of explaining "criticality" abstractly, teachers can provide contrasting paragraphs and ask students to discuss which makes a stronger analytical move. Instead of giving all feedback themselves, teachers can structure peer comparison so that students encounter multiple ways of meeting the task. These activities generate internal feedback before the final submission, when it is still actionable.

Internal feedback also supports equity when designed carefully. Students from more privileged educational backgrounds may already know how to compare their work with standards because they have prior access to academic discourse. Other students may not. Structured comparison makes this hidden practice public. It shows students how judgement works rather than assuming they already possess it.

2.3 Co-Constructing Criteria and Cultivating Evaluative Judgement

Dialogic assessment rests on two reciprocal movements. The first is the opening of academic standards to shared interpretation: students must be invited into the meanings of quality rather than merely handed the language of criteria. The second is the gradual development of students' capacity to judge quality for themselves. These two movements are inseparable. Standards cannot become meaningful to students unless they are used, tested, discussed and applied in concrete contexts. Equally, students cannot develop evaluative judgement unless they have repeated opportunities to participate in the practices through which standards are made visible. Dialogue is therefore not an optional enhancement to assessment. It is the mechanism through which standards move from teacher-held expertise to student-internalized judgement.

1. The Problem of Transferring Standards

A persistent weakness of contemporary assessment design is the assumption that standards can be transferred by specification. The common sequence is familiar: teachers write learning outcomes, construct criteria, design rubrics and distribute these documents to students before the task. Students are then expected to understand what quality requires. When misunderstanding occurs, the institutional response is often to make the rubric more detailed, the descriptors more explicit or the marking scheme more transparent. Yet this response underestimates the nature of academic standards.

Sadler's (2010) critique is important here. He argues that students need evaluative experience and tacit knowledge to interpret criteria and use feedback effectively. Criteria such as "critical analysis," "coherent argument," "original engagement," "methodological rigour" or "theoretical sophistication" do not become meaningful because they appear in a rubric. Their meaning depends on experience of judging actual work. Students must see how such qualities appear, how they vary across levels, how they interact with one another and how they are recognized by disciplinary readers.

The problem is not that criteria are useless. Criteria provide important orientation. They can reduce arbitrariness, signal what matters and offer a shared language for discussion. The problem is that criteria do not transfer standards automatically. A novice may know the words without knowing the standard. For example, a student may understand that an essay requires "critical engagement with literature" but interpret this as adding a paragraph of disagreement after each source. Another may believe that "wide reading" means citing many texts, rather than selecting sources that advance a problem. A student may follow all formal requirements and still fail to produce a persuasive academic performance because the standard is not reducible to compliance with visible features.

Academic standards contain tacit knowledge. This does not mean they are mystical, private or immune from scrutiny. It means that they are partly embodied in practices of judgement. Experienced assessors recognize patterns of quality because they have encountered many examples, participated in disciplinary conversations and learned to make comparisons. They know, often tacitly, when evidence is merely attached to a claim and when it genuinely supports an argument; when structure is mechanical and when it reflects conceptual control; when originality is superficial novelty and when it represents meaningful insight. Such knowledge can be shared, but not by transmission alone.

Dialogue is the mechanism of transfer because it allows students to encounter standards in action. Through discussion, comparison,

questioning, justification and revision, students begin to see how criteria operate in specific cases. They learn not only what the words say but how the community uses them. Standards are therefore transferred less like information and more like participation in a practice. The student becomes capable of judgement by joining, initially with guidance, the activity of judging.

2. Exemplars and the Analytic Discussion of Quality

Exemplar-based learning is one of the most important practices for opening standards to dialogue. Exemplars make quality visible in ways that abstract criteria cannot. A rubric may state that excellent work has a clear argument, but an exemplar allows students to see how an argument is established, sustained, complicated and concluded. A descriptor may mention “integration of evidence,” but an exemplar shows how evidence is woven into reasoning rather than merely inserted.

The value of exemplars does not lie simply in giving students models to imitate. Used poorly, exemplars can produce formulaic copying, anxiety or the belief that there is one correct way to perform the task. Used dialogically, however, exemplars become shared objects of inquiry. Students compare samples of different quality, identify features that distinguish them, discuss disagreements and connect their observations to criteria. This movement from seeing to articulating is crucial. Students begin to develop a vocabulary of quality grounded in concrete examples.

To and Carless’s (2016) work is important because it emphasizes the active and dialogic use of exemplars. Students do not automatically learn standards by looking at samples. They need structured opportunities to discuss what makes one response stronger than another, to test their interpretations with peers and teachers, and to connect observations to future action. The teacher’s role is not merely to reveal the “correct” judgement, but to mediate students’ developing capacity to notice and reason.

Handley and Williams (2011) similarly show that exemplars can help students understand tacit standards, particularly when accompanied by discussion. Students often value access to examples because examples reduce uncertainty, but the deeper educational value comes from examining how quality is constituted. For instance, in a postgraduate research proposal module, students might examine three literature review excerpts. The first summarizes sources sequentially; the second groups sources thematically; the third uses literature to construct a research problem. Through guided discussion, students can see that quality does not lie merely in number of references but in the intellectual work performed by the literature. This is a standard made visible through comparative dialogue.

Exemplar work also supports what Nicol describes as internal feedback. When students compare their own work with exemplars, they generate information for themselves. They notice gaps, possibilities and alternative strategies. The teacher does not need to comment on every feature because the comparison itself becomes a feedback-generating activity. Dialogue, in this sense, includes orchestrated comparison: students are placed in relation to examples, peers, criteria and their own drafts so that they can produce increasingly refined judgements.

3. Co-Constructing Criteria

If exemplars help students see quality, co-constructing criteria helps them articulate it. Student-generated criteria, marking exercises and negotiated assessment arrangements all challenge the assumption that criteria must be fully produced by teachers before learning begins. They invite students into the interpretive work behind assessment.

Student-generated criteria can begin with exemplars. Students review several pieces of work and identify features that make them more or less effective. They then propose criteria in their own language. The teacher helps refine these into disciplinary terms, clarifies misconceptions and connects them to formal programme standards. This process does not

mean surrendering academic standards to student preference. Rather, it creates a bridge between students' emerging language and the discipline's more developed language. Criteria become not a document handed down from above but a record of shared inquiry into quality.

Marking and moderation exercises are another powerful practice. Students can be asked to apply criteria to anonymized samples, assign provisional marks and justify their decisions. They then compare judgements with peers and with the teacher's judgement. Disagreement is especially valuable because it exposes the interpretive nature of assessment. Students learn that judgement is not a mechanical act of matching features to descriptors. They must weigh evidence, consider the whole performance, distinguish minor weakness from major flaw and justify standards in relation to purpose.

Such exercises also demystify moderation. In many universities, moderation is a hidden professional process: markers compare grades, discuss borderline cases and negotiate consistency. Bringing a simplified version of this process into teaching allows students to see how standards are socially maintained. They learn that academic judgement involves reasoned deliberation, not arbitrary preference. This can strengthen trust because students encounter assessment as an accountable practice rather than an opaque verdict.

Negotiated assessment contracts represent a more formal version of co-construction. In some settings, students and teachers may agree on task focus, evidence types, weightings, criteria or modes of demonstration. These contracts are especially relevant in project-based, professional, creative or postgraduate contexts where learning outcomes may be shared but pathways vary. The risk is that negotiation becomes procedural rather than educational, or that confident students gain more influence than quieter peers. For this reason, negotiated assessment must be carefully bounded. Core standards should remain clear, equity must be protected and negotiation should concern how students demonstrate learning, not whether essential learning requirements apply.

The common thread across these practices is that criteria are treated as objects for dialogue. Students do not merely receive them; they work with them. They translate, apply, question, test and revise them. In doing so, they begin to understand that standards are not simply written statements but community practices of judgement.

4. Evaluative Judgement as the Central Capability

The long-term aim of these practices is evaluative judgement. Tai, Ajjawi, Boud, Dawson and Panadero (2018, p. 467) define evaluative judgement as “the capability to make decisions about the quality of work of oneself and others.” This definition is central to the argument of the book. It shifts the aim of assessment from producing grades to developing learners who can judge quality beyond the immediate task.

Evaluative judgement includes several capacities. Students must understand what quality means in a given context. They must compare performances against standards. They must recognize strengths and weaknesses. They must justify judgements with reasons. They must decide what action follows from judgement. They must also learn that standards vary across purposes and communities. What counts as quality in a reflective journal is not identical to what counts as quality in a laboratory report, policy brief, studio portfolio or doctoral thesis.

This capability is central to higher education because graduates will continue to work in contexts where teachers are absent. They will need to decide whether an argument is strong enough, whether evidence is adequate, whether professional performance meets a standard, whether feedback from others is useful, and whether their own work requires revision. Higher education cannot prepare students for such futures by keeping judgement located solely with teachers. It must create opportunities for students to practise judgement while expert support is still available.

Dialogic assessment serves evaluative judgement by making judgement visible, discussable and repeatable. When students compare exemplars,

they practise noticing quality. When they peer review, they practise articulating judgement. When they self-assess, they practise applying standards to their own work. When they discuss criteria, they practise translating standards into usable language. When they revise after feedback, they practise acting on judgement. Each practice contributes to the gradual formation of evaluative capacity.

5. Self- and Peer Assessment as Dialogic Practice

Self- and peer assessment are often misunderstood as grading shortcuts. Institutions sometimes adopt them to reduce marking workload or to produce additional marks without increasing staff labour. This instrumental use misses their deeper educational purpose. Self- and peer assessment matter because they require students to participate in judgement.

Self-assessment asks students to compare their work against standards and make claims about quality. At its weakest, it becomes a perfunctory rating exercise: students assign themselves a mark, often strategically or inaccurately. At its strongest, it becomes a structured act of metacognition. Students identify evidence of achievement, recognize limitations, explain decisions and plan improvement. The value lies not primarily in whether the student's mark matches the teacher's mark, although calibration matters. The value lies in the reasoning process through which students learn to monitor quality.

Panadero, Jonsson and Botella's (2017) meta-analyses connect self-assessment with self-regulated learning and self-efficacy. Self-assessment supports planning, monitoring and evaluation. Students become more capable of asking: What is the goal? What standard am I aiming for? How does my current work compare? What strategy should I change? Such questions are not peripheral to learning; they are the structure of independent academic activity.

Peer assessment adds a social and dialogic dimension. Nicol, Thomson and Breslin's (2014) study of students' experiences of peer review is

especially significant because it identifies learning benefits from producing feedback as well as receiving it. When students review peer work, they must apply criteria, identify quality features, formulate reasons and communicate suggestions. This process generates reflection on their own work. In judging another's draft, students often notice issues they had missed in their own. Peer review therefore functions as a route to internal feedback.

Peer assessment also develops calibration. Students compare their judgements with those of peers and teachers. They encounter disagreement and must decide whether the disagreement reflects misunderstanding, alternative interpretation or genuine ambiguity in the standard. This is precisely the kind of judgement work that expert assessors perform. When designed well, peer assessment inducts students into a community of evaluators.

However, self- and peer assessment require careful design. Students need criteria, exemplars, modelling, practice and trust. Without these, peer comments may be superficial, overly harsh, overly generous or ignored. Students may distrust peers' authority, fear damaging relationships, or lack confidence in their own judgement. Dialogic structures can address these risks: training students to give feedback, using feedback-quality criteria, allowing clarification, requiring response plans and ensuring that teacher judgement remains available as a reference point.

The aim is not to make students substitute for teachers. It is to help students become participants in judgement. Teacher expertise remains crucial, but its role shifts. The teacher designs the conditions, models evaluative reasoning, calibrates standards and supports reflection. Students increasingly take responsibility for judging quality.

2.4 Designing for Dialogue: Conditions, Constraints, and Trade-offs

The case for dialogic assessment can easily be misunderstood as a moral appeal: teachers should talk more with students, students should participate more actively, and assessment should become more humane. Such claims are not wrong, but they are insufficient. Dialogue is not sustained by goodwill alone. It requires curriculum structures, time, staffing, assessment sequencing, institutional legitimacy, relational continuity and defensible standards. Without these enabling conditions, dialogic assessment becomes either an individual teacher's heroic labour or a rhetorical aspiration attached to systems that remain fundamentally monologic.

This section therefore treats dialogic assessment as a design problem under constraint. The question is not whether dialogue is desirable in the abstract, but where and how it can be built into higher education assessment systems marked by large cohorts, modular curricula, casualized labour, quality assurance demands, digital platforms and finite resources. A credible account of dialogic assessment must acknowledge that not every assessment moment can be dialogic in the same way. It must also specify the trade-offs involved when institutions invest in dialogue at some points and not others.

1. Structural Conditions: Loops, Sequences and Programme-Level Design

The first enabling condition is structural: assessment must be sequenced so that feedback can be used. Dialogue requires loops. A loop connects work, feedback, interpretation and reuse. The student produces work; receives or generates information about its quality; makes sense of that information through comparison, discussion or reflection; and then applies it to a revised or subsequent performance. Without this loop, feedback remains largely explanatory. It may justify a mark, but it cannot reliably support learning.

Many higher education assessment systems fail at precisely this point. A student submits an essay at the end of a module, receives comments after the module has finished, and then moves to a different course with a different teacher, task and standard. The feedback may be thoughtful, but the curriculum offers no structured occasion for reuse. Students are expected to transfer insights independently across fragmented assessment events. This is a weak design for dialogue because it treats uptake as an individual responsibility rather than a curriculum responsibility.

Dialogic assessment therefore requires nested or programmatic assessment design. Tasks should be connected over time so that early work feeds later work. A research proposal should lead into a literature review; a literature review into a project report; a project report into a reflective commentary or presentation. A first essay should not merely be marked and returned; it should generate specific learning that students are required to demonstrate in a second essay. A portfolio should contain evidence of development, not only final products. In such designs, feedback becomes consequential because students must use it.

Nested designs also allow standards to be revisited. Students encounter criteria not once, at the moment of assignment release, but repeatedly through exemplars, peer review, draft work, feedback, revision and later application. This repetition is essential because understanding standards is developmental. A single rubric briefing rarely alters judgement. A sequence of tasks can.

Programme-level design extends this principle beyond individual modules. If each module operates as a self-contained assessment island, students may experience feedback as episodic and non-transferable. A programme-level architecture can identify key capabilities—argumentation, evidence use, methodological reasoning, reflective judgement and professional communication—and design repeated opportunities for students to practise and receive feedback on them. Dialogue then becomes cumulative rather than accidental.

Such designs require timetable and workload models that make conversation feasible. A curriculum cannot claim to value dialogue if all marking time is allocated to final grading and no time is allocated to pre-submission discussion, exemplar analysis, peer moderation or feedback uptake activities. Dialogic assessment requires institutional recognition that feedback conversations, calibration exercises and assessment literacy activities are teaching work, not optional extras.

2. Relational Conditions: Trust, Safety and Continuity

The second enabling condition is relational. Students do not engage with assessment dialogue simply because an opportunity exists. They engage when they trust that the exchange is educationally safe, meaningful and worth the risk. Assessment is emotionally charged because it evaluates work that students often experience as an extension of competence, identity and belonging. Dialogue about assessment therefore requires psychological safety: students must be able to ask naïve questions, disclose uncertainty, test interpretations and receive critique without feeling humiliated.

Trust is built over time. It is difficult to establish in systems where students encounter teachers only briefly, where marking is anonymous and detached from teaching, or where feedback is returned by casual staff with whom students have no continuing relationship. Anonymous marking can support fairness by reducing some forms of bias, but it can also weaken the relational conditions through which feedback becomes meaningful. A marker who does not know the student's prior work, intentions or developmental trajectory may produce accurate comments but cannot easily enter a sustained dialogue about growth.

Casualized teaching intensifies this problem. When tutors are employed on short-term or fragmented contracts, they may have limited time for feedback discussion, limited access to programme-level assessment conversations and limited continuity with students. They may mark work without having designed the task, taught the module or

participated fully in standard-setting. This does not mean casualized teachers lack expertise; rather, the employment structure often prevents them from exercising that expertise dialogically. Dialogue requires continuity, and continuity requires institutional conditions.

Relational continuity is especially important for students who are less confident in academic culture. Students from non-traditional backgrounds, first-generation students, international students, disabled students, students writing in an additional language and students who have previously experienced assessment failure may be less likely to seek clarification or challenge feedback. If dialogue is available only to those who ask for it, it may reproduce inequality. A dialogic assessment system must therefore embed opportunities for all students, not only the confident or well-connected.

The relational condition does not require the abolition of critique. Indeed, trust enables more demanding critique because students are more likely to hear challenge as part of development rather than exclusion. The aim is not comfort, but sustainable engagement with difficulty. Students need to experience assessment dialogue as a space in which standards are serious, but their developing capacity to meet those standards is taken seriously too.

3. The Scale Problem: Dialogue with Hundreds

The most common objection to dialogic assessment is scale. A teacher with a cohort of twenty can hold individual conferences, read drafts and discuss feedback in depth. A lecturer with three hundred students cannot do this in the same way. The scale problem is real, and any model that ignores it should not be taken seriously.

However, scale does not make dialogue impossible. It changes its form. Large-cohort dialogic assessment requires distributed dialogue rather than universal one-to-one dialogue. Peer dialogue is one of the most important mechanisms. Students can review one another's drafts, compare exemplars, discuss criteria and generate feedback in structured groups.

This does not replace teacher expertise, but it extends the space in which students practise judgement. For peer dialogue to work, it must be scaffolded: students need criteria, examples of useful feedback, calibration tasks and opportunities to reflect on feedback received.

Group feedback is another scalable practice. Instead of writing the same comment repeatedly on individual scripts, teachers can identify common patterns and discuss them with the cohort. This can be highly dialogic if students are asked to compare these patterns with their own work, annotate examples, ask questions and create action plans. Group feedback may have limited value when students cannot relate it to their work or act on it.

Technology-mediated dialogue also has potential. Discussion boards, annotated exemplars, collaborative documents, audio feedback, screencasts, polling tools, learning analytics and peer-review platforms can widen participation and allow asynchronous engagement. Students who are reluctant to speak in class may ask questions online. Teachers can respond to recurring misunderstandings publicly. Peers can comment on drafts without needing to be physically present. Digital tools can preserve traces of dialogue for later use.

Yet technology does not solve the problem by itself. It can easily reproduce transmission at scale: automated comments, dashboard alerts, generic annotations and discussion forums with little meaningful interaction. Technology supports dialogue only when it is designed around interpretation, comparison, response and reuse. The question is not whether feedback is digital, but whether the digital environment enables students to make sense of standards and act on feedback.

Selective high-touch points are also necessary. Not every assessment moment requires individual dialogue, but some do. Programmes should identify moments where teacher-student conversation has the highest developmental value: transition into first-year academic writing, major project proposals, dissertation planning, professional practice preparation, or feedback after a significant failure. Concentrating

intensive dialogue at these points may be more effective than spreading thin comments across every task.

Some things do not scale well. Extended individual tutorials for every assignment, detailed comments on multiple drafts from every student, and open-ended negotiation of criteria in very large cohorts are rarely sustainable without major staffing investment. A realistic dialogic architecture must therefore combine high-touch, medium-touch and low-touch practices. The aim is not to make every interaction intensive, but to ensure that every student encounters meaningful opportunities for uptake.

4. The Standardization Tension

Dialogic assessment also faces a standardization tension. Higher education institutions must ensure consistency, comparability and fairness across markers, modules and cohorts. Students should not receive different standards depending on which tutor they happen to meet. Programmes must be able to defend grades to external examiners, professional bodies and appeals processes. Dialogue and negotiated standards can appear to threaten this consistency.

The tension is genuine but not fatal. It depends on what is being negotiated. Dialogic assessment should not imply that essential standards are infinitely flexible or that students can bargain over academic quality. Rather, dialogue concerns the interpretation and application of standards in context. The formal standard may remain stable, while its meaning is explored through examples, discussion and student work. For instance, a criterion such as “critical engagement with literature” can remain a programme standard, but students and teachers can examine how it appears differently in essays, reviews, research proposals and dissertations.

Consistency does not require the elimination of judgement. In complex assessment, consistency is achieved through shared interpretive practices as much as through written criteria. Marker calibration, moderation discussions, shared exemplars and collective review of

borderline cases are themselves dialogic practices. They demonstrate that standards are maintained through professional conversation, not merely through documents. Bringing students into appropriate parts of this conversation can strengthen rather than weaken fairness, because students gain access to how standards are interpreted.

Quality assurance systems often prefer written evidence because it is auditable. Dialogue, especially oral dialogue, can appear ephemeral. The design challenge is to preserve the educational value of dialogue while producing appropriate records. This can be done through feedback action plans, student response sheets, portfolio reflections, peer-review logs, revision commentaries and moderation notes. Such records should not replace dialogue, but they can document its outcomes.

There is also a fairness issue in negotiated assessment. If some students receive more dialogue than others, or if negotiation is informal, advantages may accumulate unevenly. Dialogic opportunities must therefore be transparent and structured. Students should know what kinds of questions they may ask, what kinds of choices are negotiable, what standards remain fixed and how feedback conversations relate to grading. Dialogue must be available as part of the curriculum, not as private access to the assessor.

5. Trade-Offs and Design Heuristics

Because resources are finite, dialogic assessment requires decisions about where to invest effort. The following heuristics offer a practical framework.

First, invest in dialogue where feedback can be reused. A short conversation before a task or between draft and final submission is usually more valuable than extensive comments after a terminal assessment. If students cannot act on feedback, the educational return is limited.

Second, invest in dialogue where standards are tacit or complex. Tasks involving judgement, originality, professional reasoning, critical analysis, design, reflection or research require more dialogic support than tasks

with straightforward right-or-wrong answers. The more complex the standard is, the more students need examples, discussion and calibration.

Third, invest dialogue at transition points. Students need particular support when entering university, moving into disciplinary specialization, beginning independent research, entering placement, or preparing capstone work. These moments often involve new standards and new identities.

Fourth, invest dialogue where failure has high consequences. Students at risk of failing, repeating or leaving may need targeted conversations that help them interpret what has happened and what action is possible. Generic written feedback is rarely sufficient in such cases.

Fifth, use peer and group dialogue to widen access, but preserve teacher judgement for calibration. Students benefit from judging and discussing one another's work, but teacher expertise remains necessary to prevent misconceptions from circulating unchecked. The teacher need not comment on everything but must design and anchor the evaluative environment.

Sixth, document the outcomes of dialogue without reducing dialogue to paperwork. A short student action plan may be more useful than a long teacher comment. A revision note may show feedback uptake more clearly than a feedback sheet. Records should support learning and accountability simultaneously.

The major risk is partial implementation. Institutions may add dialogic activities at the margins—an optional feedback workshop, a peer-review exercise, a reflective coversheet—while leaving the high-stakes assessment core unchanged. In such cases, students quickly learn what really counts. If final grades are still determined by isolated, monologic tasks, dialogue becomes supplementary rather than constitutive. It may improve student experience but will not transform the assessment ecology.



Dialogue in Practice

Illustrative Designs for Assessment Reconceived

3.1 The Feedback Conversation: Dialogic Feedback in Disciplinary Writing

The examples in this chapter are illustrative designs informed by the literature, not reports of original case-study research. Writing-intensive disciplines provide one of the clearest sites for examining dialogic assessment in practice. In the humanities and social sciences, student achievement is often demonstrated through essays, literature reviews, research proposals, critical commentaries, policy analyses, reflective papers and dissertations. These forms of work are not merely containers for knowledge already acquired. They are themselves modes of disciplinary thinking. To write an essay in sociology, history, literature, education or politics is to make an argument, position evidence, interpret concepts, engage with traditions of debate and adopt the voice of a field. Assessment of writing is therefore assessment of disciplinary participation.

This makes feedback on writing especially consequential and especially difficult. Teachers often know when an essay lacks argument, when evidence has been used mechanically, when a concept has been named but not understood, or when a paragraph has failed to do intellectual work. Yet the comments used to express these judgements “develop the argument,” “engage more critically,” “avoid description,” “strengthen your analysis,” “clarify the conceptual frame” are frequently opaque to students. The problem is not simply that students ignore feedback. It is

that disciplinary writing standards are partly tacit, embedded in ways of reading and judging that are learned through participation. For this reason, writing-based assessment offers a particularly important case for the book's central claim: feedback becomes educative when it is organized as dialogue around quality, not merely delivered as comment after judgement.

1. The Baseline Problem: The Comment-and-Grade Cycle

The conventional feedback cycle in essay-based modules is familiar. Students attend lectures and seminars, receive an assignment brief and rubric, write an essay independently, submit it near the end of the module, and later receive a grade with written comments. The comments may be thoughtful and detailed. They may identify strengths, diagnose weaknesses and suggest improvements. Yet the cycle is structurally weak. By the time feedback arrives, the assessed work is complete, the emotional meaning of the grade has already dominated attention, and the next assignment may belong to a different module, topic, genre or teacher.

This “comment-and-grade” model rests on a fragile assumption: that students will read comments, understand their disciplinary meaning, translate them into strategies, and transfer those strategies to a different future task. In practice, each step is uncertain. Some students do not read feedback closely, especially if the grade meets expectations. Others read it but do not understand what it requires. Others understand the general direction but cannot see what concrete action follows. A student told to “use theory more analytically” may not know whether this means defining more concepts, applying one concept more deeply, comparing theoretical perspectives, or using theory to reshape the research question. A student told that an essay is “too descriptive” may remove contextual detail but still fail to construct analysis.

The discontinuity between assignments intensifies the problem. In a modular curriculum, an essay on Weber in one semester may be followed by a research methods report, a policy brief or a group presentation. The

next task may not invite obvious reuse of prior feedback. Students are expected to abstract general lessons from one genre and carry them into another. Some experienced students can do this; many cannot. Feedback thus becomes a record of past performance rather than a usable bridge to future learning.

There is also a relational difficulty. Written comments often arrive as a one-way judgement from an absent assessor. The student cannot easily ask what a phrase means, explain the intention behind a decision, or test whether a proposed revision would improve the work. The teacher cannot discover whether the student misunderstood the task, lacked confidence, misread the rubric, or made a strategic decision under time pressure. The feedback event therefore fails to become a feedback conversation. It remains an annotated verdict.

This baseline pattern is not the result of teacher indifference. It is produced by assessment architecture. The curriculum gives teachers few opportunities to enter the writing process before submission and gives students few incentives or occasions to use feedback afterward. The task is assessed as a finished product, but the capability students need—the capacity to think and write within a discipline—develops through iterative, interpretive activity. The redesign of writing feedback must therefore intervene in the sequence itself.

2. Redesigning the Writing Feedback Cycle

A dialogic redesign begins by moving feedback earlier, making it usable, and positioning students as participants in the conversation. The central shift is from “submit–mark–return” to “draft–discuss–revise–reflect.” Two-stage or “feedback-then-revise” assignments are the clearest example. Students submit a draft or partial version of the work, receive feedback from peers, tutors or both, and then revise for a final submission. The final version may be accompanied by a commentary explaining what feedback was used, what was rejected and why.

The educational power of two-stage design lies in its alteration of responsibility. Students cannot treat feedback as an after-the-fact explanation of a grade. They must decide what to do with it. A comment such as “the literature review is organized source by source rather than around a problem” becomes a prompt for revision. The student must reorganize, compare alternatives and judge whether the new structure better supports the argument. Feedback becomes part of composing, not merely part of grading.

Draft tutorials add a more explicitly dialogic element. In a tutorial, the student can explain what they are trying to do and where they are uncertain. The teacher can respond not only to the text but to the student’s intention. This matters because writing often contains unrealized possibilities. A draft may appear weak because the student has not yet found a structure, but the underlying idea may be promising. A written comment might simply identify lack of coherence. A tutorial can ask: What is the central problem? Which paragraph best expresses it? What would happen if the essay began there? In such exchanges, feedback becomes jointly diagnostic.

Feedback request sheets are another important mechanism. Before submitting a draft or final assignment, students are asked to identify the aspects of the work on which they most want feedback. They might answer questions such as: What were you trying to achieve in this essay? Which part are you least confident about? Which previous feedback did you try to use? What kind of response would be most helpful at this stage? These sheets shift the agenda. The student is no longer merely exposed to the teacher’s judgement; the student directs attention to problems they are already beginning to recognize.

Feedback request sheets also reveal students’ developing evaluative judgement. A novice may ask only, “Is this good enough?” A more developed student may ask, “Does the theoretical framework actually organize the analysis, or does it sit separately in the introduction?” This second question shows that the student is learning to think like a

disciplinary writer. The teacher's response can then meet the student at a higher level of judgement.

Dialogue around exemplars before submission is equally important. Students often receive rubrics but lack experience applying them. Exemplar discussion makes quality visible. In an essay-based social science module, students might compare three introductions: one that merely announces the topic, one that summarizes the debate, and one that constructs a problem and establishes a position. The class can discuss which introduction creates the strongest intellectual momentum and why. Students then return to their own drafts and diagnose their openings. The exemplar becomes a shared reference point for feedback before the assignment is submitted.

These redesigns are not discrete techniques. They are parts of an assessment architecture. Exemplars help students anticipate quality; request sheets invite students to articulate their needs; draft tutorials mediate interpretation; revision requires action; reflective commentaries make uptake visible. Together they convert feedback from a terminal comment into an ongoing dialogue about writing.

3. Mechanisms: Feed-Forward, Sense-Making and Disciplinary Knowing

The first mechanism at work is feed-forward across linked tasks. In writing-intensive disciplines, improvement rarely occurs through isolated correction. Students develop through repeated engagement with similar but increasingly demanding writing problems: constructing claims, positioning sources, moving from description to analysis, integrating theory, organizing evidence and developing voice. Dialogic feedback supports this development when tasks are linked so that feedback from one stage is required in the next. A proposal feeds a literature review; a literature review feeds an essay; an essay feeds a presentation or reflection; a reflection feeds a later research design. The loop makes feedback consequential.

The second mechanism is the shift of cognitive work onto the student as sense-maker. In the comment-and-grade model, the teacher does most of the visible evaluative labour. The teacher diagnoses the problem, writes the comment and suggests improvement. The student may remain passive. In dialogic designs, students must interpret, compare, decide and act. They must judge whether feedback is relevant, which suggestions matter most, how different comments relate, and how revision should proceed. This does not reduce teacher responsibility; it redistributes cognitive activity so that students practise the judgement the discipline requires.

The third mechanism is the building of disciplinary ways of knowing. Writing feedback in the humanities and social sciences is not simply about improving sentences. It is about learning what counts as an argument, evidence, explanation and interpretation within a field. A historian's use of evidence differs from a sociologist's; a literary critic's interpretation differs from a policy analyst's; an education researcher's treatment of theory differs from a reflective practitioner's. Dialogue makes these differences discussable. Students learn not only that quality matters but how quality is recognized in a particular discipline.

This is why disciplinary writing feedback cannot be reduced to generic study skills. A comment such as "be more critical" does not have one universal meaning. In sociology, it may mean interrogating assumptions, situating claims within theoretical traditions or showing how evidence complicates common explanations. In literature, it may mean developing an interpretation attentive to language, form and context. In politics, it may mean weighing competing explanations, institutional constraints or normative implications. Dialogic feedback allows teachers to unpack these disciplinary meanings in relation to actual student texts.

The fourth mechanism is comparison. Students generate internal feedback when they compare their work with exemplars, peer drafts, criteria, previous attempts and teacher responses. In writing, comparison is especially powerful because quality is relational and often hard to define

abstractly. Students can see that one paragraph quotes a theorist and then moves on, while another uses the theorist to transform the analysis. They can see that one conclusion repeats the introduction, while another clarifies what the argument has shown. These comparisons develop the student's eye for quality.

4. Evidence and Outcomes

The research literature offers support for dialogic feedback in writing, but its scope must be treated carefully. For example, Hill and West (2022) traced four students through a dialogic feed-forward design and reported benefits alongside implementation challenges. Such context-specific evidence helps explain how feedback can support learning; it does not establish that dialogue automatically improves writing outcomes across settings.

Nicol's work is frequently invoked as an important conceptual shift because it emphasizes feedback as a process that should support self-regulation and dialogue, including teacher-student and peer discussion. In writing-based contexts, this means that students must learn to use feedback to monitor their own composing decisions. Beaumont, O'Doherty and Shannon (2011) similarly argue for reconceptualizing feedback as a dialogic cycle rather than one-way transmission. Their contribution is important because it addresses a recurring weakness in higher education: students experience feedback as something done to them after the work is finished, rather than as part of learning how to produce better work.

The accounts analysed by Hill and West (2022) illustrate several possible benefits. Students may find that dialogue helps them understand comments that would otherwise remain unclear. In draft conferences or supervisory discussions, they can ask what a comment means, check whether they have interpreted it correctly, and identify specific revision actions. Dialogic feedback can reduce misunderstanding, particularly

where written comments are brief or where students and teachers have different expectations about the purpose of feedback.

The design also creates opportunities for engagement. When students must bring questions, request feedback or revise in response, they have a reason to treat feedback as usable information. The feedback request sheet is especially valuable in this respect because it requires students to enter the assessment conversation before receiving teacher judgement. It can reveal whether students are monitoring their own writing and whether they are beginning to recognize disciplinary problems.

The quality of revision can also improve, especially when feedback is timely, specific and connected to a required revision. Students may reorganize arguments, clarify claims, improve use of sources and strengthen introductions or conclusions. However, reported benefits from a particular design do not establish that adding a conference to written feedback will improve revision scores. The quality of the dialogue, the student's readiness, the nature of the task and the assessment design all matter.

Dialogic feedback can also support feedback literacy. Students learn what feedback is for, how to ask for it, how to interpret it, how to manage critique and how to act. These capacities are not secondary outcomes; they are central to disciplinary writing development. A student who can ask a precise question about argument structure is already demonstrating growth as a writer. A student who can explain why they rejected one piece of feedback and acted on another is practising evaluative judgement.

The affective outcomes are mixed but significant. Writing feedback can produce anxiety, shame, defensiveness and loss of confidence because writing exposes thinking. Dialogic feedback can reduce some of these risks by allowing clarification and humanizing the assessor. Students may experience motivation when feedback becomes a conversation rather than a verdict. At the same time, dialogue can also heighten vulnerability. A draft tutorial requires the student to discuss unfinished work; peer review

exposes writing to classmates; critique may feel personal. The relational framing of feedback is therefore crucial.

Workload remains a major constraint. Draft tutorials, two-stage marking and detailed revision feedback require time. Teachers may find that dialogic processes increase front-end labour even if they reduce repeated commenting later. Some students also disengage. They may fail to submit drafts, ignore peer review, attend tutorials unprepared, or treat revision mechanically. Dialogic design cannot guarantee participation. It can only make participation more likely and more meaningful by embedding it into the assessment structure.

5. Disciplinary Specificity: Writing-Intensive Fields and What Transfers

Dialogic feedback in writing-intensive fields differs from feedback in problem-based or quantitative fields because the object of judgement is often holistic, interpretive and rhetorically situated. In a mathematics problem set, feedback may identify an incorrect step, a flawed assumption or an inefficient method. Although conceptual dialogue can be important, the correctness of the solution may be more determinate. In an essay, the problem is rarely only correctness. The quality of the work depends on framing, selection, emphasis, interpretation, evidence, structure and voice.

This does not mean that quantitative fields are non-dialogic or that writing fields lack rigour. It means that the tacit dimensions of quality are especially visible in writing. A strong essay is not produced by adding all required components. Its quality lies in how those components interact. The introduction creates a problem; the literature review positions the debate; the theory guides interpretation; the evidence supports and complicates claims; the conclusion clarifies significance. Feedback must therefore address the whole intellectual performance.

Shulman's (2005) concept of signature pedagogies is useful here. Disciplines have characteristic ways of teaching students to think, perform

and value. In writing-intensive disciplines, seminars, tutorials, peer critique, essay supervision and dissertation meetings can be examined through this lens because they induct students into disciplinary judgement through talk about texts. In studio and creative disciplines, critique serves a similar function. In clinical fields, case discussion and supervised practice may play this role. The transferable principle is not a particular feedback format, but the idea that students learn quality through guided participation in the discipline's evaluative practices.

What transfers across fields is the need for comparison, dialogue, uptake and judgement. Students in quantitative fields also benefit from explaining reasoning, comparing solution strategies, reviewing worked examples and reflecting on errors. Students in professional fields benefit from feedback conversations around performance evidence. The difference lies in the nature of the standards and artifacts. Dialogic assessment must therefore be discipline-sensitive, not generic.

3.2 Co-Constructed and Negotiated Assessment: Students as Partners in Judgement

The preceding section examined dialogic feedback in the teaching of disciplinary writing, where students enter a conversation about quality through drafts, exemplars, feedback requests and revision. This section moves one step further. It examines cases in which students do not only respond to assessment but participate in shaping the assessment itself. These practices include co-creating criteria, designing rubrics, negotiating assessment tasks, proposing topics or formats, and participating in judgement through moderation, peer review or partnership structures. They represent a more ambitious form of dialogic assessment because they alter not only the communication surrounding assessment, but the authority through which assessment is designed and interpreted.

The central claim of this section is that co-constructed assessment can make standards more intelligible, increase ownership and motivation, and

develop evaluative judgement. However, these benefits are not automatic. Student participation in assessment design can be transformative, but it can also be tokenistic, inequitable, strategically manipulated or constrained by institutional and professional requirements. The question is therefore not whether students should participate in assessment, but under what conditions participation becomes educationally meaningful rather than cosmetic.

1. Co-Creating Criteria and Rubrics

Co-creation of assessment criteria is one of the most concrete forms of dialogic assessment. Instead of presenting students with a finished rubric, teachers invite students to participate in identifying, interpreting and refining the criteria by which work will be judged. This may occur through discussion of exemplars, analysis of previous student work, collaborative drafting of descriptors, or structured comparison between student-generated criteria and formal programme standards. The purpose is not simply to make students feel included. It is to help them understand what quality means by engaging in the practice of naming and justifying it.

A common design begins with exemplars. Students are given several pieces of work at different levels of quality, sometimes anonymized from previous cohorts or constructed by the teacher. Working individually and then in groups, they identify what makes one piece stronger than another. They may attend to argument, evidence, structure, originality, clarity, method, presentation or disciplinary voice. The teacher then facilitates a discussion in which students articulate criteria in their own language. These emerging criteria are compared with the teacher's or programme's criteria, and the final rubric is produced through a guided process of refinement.

This process is educationally significant because it reverses the usual sequence. In conventional criteria-based assessment, students receive the language of standards before they have participated in the judgement that

gives that language meaning. In co-creation, students first encounter concrete performances and then develop criteria from comparative judgement. This makes criteria less abstract. Students begin to see that “critical analysis” is not a phrase in a rubric but a recognizable quality in a piece of work; that “coherence” is not merely tidy structure but the relation between purpose, evidence and argument; that “originality” does not mean eccentricity but a meaningful contribution within constraints.

Co-created rubrics can also increase clarity of expectations. When students help formulate criteria, they must ask what the task is for, what kind of evidence matters, what counts as achievement and how levels of quality differ. Such questions are rarely asked when rubrics are simply distributed. The co-creation process therefore functions as assessment literacy work. It teaches students how assessment is constructed and why certain features of work are valued.

The strongest versions of co-created criteria do not end with the production of a rubric. They are followed by marking and moderation workshops. Students apply the criteria to sample work, compare marks, justify decisions and discuss disagreements. These activities are crucial because criteria are only fully understood in use. A group may agree that “evidence” matters, but disagree about whether a particular paragraph uses evidence effectively. That disagreement becomes a site of learning. Students discover that judgement involves weighing, interpretation and justification rather than mechanical application.

Moderation workshops also expose the social nature of academic standards. Students often imagine marking as an individual act of preference or authority. When they participate in moderated judgement, they see that standards are maintained through discussion, comparison and reason-giving. They also see that assessors may disagree in good faith. This can strengthen trust, provided the teacher frames disagreement not as inconsistency but as part of the interpretive work of assessing complex performances.

Co-creating criteria is intended to increase student ownership, self-awareness and influence within the assessment process. Discussing how criteria are constructed can help students understand expectations and see why the task matters. These are plausible benefits, not guaranteed outcomes: co-creation does not by itself establish improved grades, deeper learning or equitable participation. Outcomes depend on facilitation, scaffolding, task design and the surrounding assessment culture.

2. Negotiated Assessment and Learning Contracts

Negotiated assessment has a longer history than contemporary students-as-partners discourse sometimes acknowledges. It is connected to traditions of adult learning, learner autonomy, experiential education and curriculum negotiation. In negotiated assessments, students participate in deciding aspects of the assessment task: the topic, product, evidence base, weighting, format, timeline or criteria. Learning contracts are one of the classic forms of this tradition. As described by Anderson, Boud and Sampson (1996), learning contracts invite students to define learning objectives, identify resources, propose evidence of achievement and agree on assessment procedures within a framework.

The logic of learning contracts is that mature learners should not be treated as passive recipients of pre-designed tasks. They can participate in planning and evidencing their own learning. This is particularly relevant in professional education, work-based learning, independent study, postgraduate research, project-based modules and interdisciplinary programmes. In such contexts, a single standardized assignment may not capture the diversity of student goals or professional trajectories. A negotiated task can allow students to align assessment with meaningful problems, contexts and identities.

For example, in a postgraduate education module, students might all be required to demonstrate understanding of curriculum theory, but they may negotiate the form of evidence: a critical essay, a curriculum design portfolio, a policy analysis, or an action-research proposal. In a social

science research module, students might select topics relevant to their own interests while still meeting common standards for research design, ethics, literature engagement and methodological justification. In a professional programme, students might negotiate evidence from placement or practice while remaining accountable to accreditation requirements.

Negotiated assessment can support motivation because it increases relevance. Students are more likely to invest effort when they see the task as connected to their interests, identities or future practice. It can also support self-regulation because students must plan their work, identify standards, manage timelines and justify decisions. The assessment process becomes not only a test of content knowledge but an exercise in educational agency.

However, negotiation requires boundaries. Not everything can be negotiated. Core learning outcomes, professional competencies, safety requirements, academic integrity, and essential disciplinary standards must remain stable. The negotiation should concern how students demonstrate learning, not whether the learning matters. Without clear boundaries, negotiated assessment can become confusing, inequitable or vulnerable to gaming. Students may select easier formats, avoid challenging outcomes, or negotiate strategically for advantage. Teachers must therefore distinguish legitimate flexibility from erosion of standards.

Learning contracts also require careful support. Students may not initially know how to propose a viable assessment task or evidence achievement. Without scaffolding, negotiated assessment can favour confident students who understand academic expectations and can advocate for themselves. Less confident students may choose conventional tasks even when alternatives would better support their learning. Thus, negotiation must be facilitated through examples, templates, staged proposals, peer discussion and teacher feedback.

Contemporary revivals of negotiated assessment are often framed through student agency, personalization, decolonizing curricula, inclusive assessment and authentic learning. These revivals are valuable, but they

must avoid romanticizing choice. Choice is not automatically empowering. It becomes educational when students are supported to make informed, standards-aware decisions and to reflect on the consequences of those decisions. Negotiated assessment is therefore best understood as structured autonomy, not unlimited freedom.

3. Students-as-Partners and the Redistribution of Authority

Co-constructed assessment is closely related to the students-as-partners movement. The work of Cook-Sather, Bovill and Felten (2014), and Healey, Flint and Harrington (2014), conceptualizes partnership as a relationship in which students and staff collaborate in teaching, learning, curriculum and inquiry. Partnership is not the same as student satisfaction, consultation or consumer feedback. It rests on reciprocity, respect, shared responsibility and the recognition that students bring expertise from their lived experience of learning.

Applied to assessment, the students-as-partners framing is significant because assessment is one of the most power-laden areas of higher education. Teachers define tasks, set standards, interpret criteria and award grades. Students are usually positioned as those who must comply with requirements and accept judgement. Partnerships unsettle this arrangement by asking where students can participate in designing tasks, interpreting criteria, reviewing assessment processes, generating feedback or evaluating the fairness and meaning of assessment.

This redistribution of authority is ethical as well as pedagogical. It recognizes students as agents in their own education. It also acknowledges that assessment systems affect students' identities, opportunities and sense of belonging. If students are excluded from all assessment design, then assessment remains something done to them. Partnership seeks to make it something done with them, within appropriate limits.

Voice is central to this movement, but voice must be understood carefully. Student voice is not merely the collection of opinions after a module has ended. Nor is it reducible to survey data. Genuine voice

implies access to decision-making, influence over processes and participation in shaping meanings. In assessment, this may involve students helping to design criteria, reviewing the clarity of assignment briefs, co-creating exemplars, participating in assessment panels, contributing to programme-level assessment review or collaborating on inclusive assessment redesign.

The ethical challenge is that partnership can easily become symbolic. Institutions may celebrate student voice while leaving authority untouched. Students may be invited to comment on assessment after major decisions have already been made. A small group of high-achieving students may be selected as partners and treated as representative of all. Participation may become a badge of institutional progress rather than a redistribution of power. For this reason, partnership in assessment must be evaluated not by the language used but by the influence students have.

4. Mechanisms: Ownership, Tacit Standards and Evaluative Judgement

The benefits of co-constructed assessment can be explained through four main mechanisms: ownership, the surfacing of tacit standards, metacognition and evaluative judgement.

First, co-construction can increase ownership. When students help shape criteria or tasks, they may experience assessment as more meaningful and less arbitrary. They understand why the task has been designed as it has, what the criteria are intended to capture and how their own choices relate to learning outcomes. Ownership can increase motivation because students see themselves as participants in the assessment process rather than subjects of it.

Second, negotiation surfaces tacit standards. Much assessment knowledge remains implicit in teacher judgement. When students ask why a criterion matters, why one format is acceptable and another is not, or why a proposed task does not meet the standard, teachers are required to articulate assumptions that may otherwise remain hidden. The

negotiation process therefore makes standards more explicit, but in a dialogic rather than merely documentary way. Standards are clarified through cases, reasons and boundaries.

Third, co-construction develops metacognition. Students must think about how they learn, what evidence demonstrates learning, what quality looks like and what strategies will help them meet standards. A student designing a learning contract must ask: What am I trying to learn? How will I know I have learned it? What evidence will persuade others? What criteria should apply? These are metacognitive questions.

Fourth, co-construction develops evaluative judgement. When students participate in criteria design, marking exercises or moderation workshops, they practise deciding what counts as quality. They learn to justify judgements, compare standards, recognize ambiguity and calibrate their interpretations against others. This connects directly to the long-term aim of dialogic assessment: students should become capable judges of their own and others' work.

The development of evaluative judgement is especially visible in co-created criteria followed by moderation. Students first propose criteria, then apply them, then revise their understanding in light of disagreements. This sequence mirrors professional judgement. Academics do not maintain standards by reading rubrics alone; they discuss cases, compare work, debate boundaries and refine shared expectations. When students participate in these activities, they enter the practice of judgement itself.

5. Tensions and Failure Modes

Co-constructed assessment is promising, but it has several predictable failure modes. The first is unequal participation. Not all students enter partnership spaces with the same confidence, language resources, cultural capital, time or sense of entitlement. Students who are articulate, high-achieving or familiar with university norms may dominate discussion. Students who are first-generation, international, disabled, working-class, racially minoritized, linguistically marginalized or anxious about authority may participate less visibly. If unaddressed, co-construction can reproduce advantage while appearing democratic.

The second failure mode is tokenism. A teacher may invite students to suggest criteria but reserve the right to ignore them without explanation. A programme may ask students for assessment preferences but make no structural changes. A rubric may include a student-generated phrase while the underlying standards remain unchanged and unexplained. Such practices create the appearance of participation without meaningful influence. They may even increase cynicism because students recognize that their voice has been used symbolically.

The third failure mode is gaming. When students participate in designing criteria or negotiating tasks, they may seek to make assessment easier, reduce workload, avoid disliked components or increase the weight of familiar formats. This is not evidence that students are morally deficient; it is a predictable response in grade-driven systems. If assessment is experienced primarily as a high-stakes competition, students will often act strategically. Dialogic assessment must therefore be designed to shift the conversation from advantage to quality, while retaining safeguards.

The fourth failure mode is confusion. Students may feel less secure when assessment is negotiated because expectations appear less fixed. Some students prefer detailed instructions and may interpret negotiation as lack of teacher authority. This is especially likely when negotiation is introduced suddenly without explanation. Teachers must therefore clarify what is fixed, what is open, and why participation matters.

The fifth failure mode is conflict with accreditation or professional standards. In professional programmes, students cannot negotiate away required competencies. A nursing, engineering, teaching or social work programme must satisfy external standards. Even in non-professional disciplines, departments must ensure comparability across cohorts and modules. Co-construction must therefore operate within frameworks. Students may help interpret, evidence or contextualize standards, but not abolish them.

The sixth failure mode is staff overload. Facilitating co-created criteria, negotiated tasks or partnership processes requires time and skill. Poorly supported staff may default to superficial participation or abandon the process as inefficient. Institutions that encourage partnership without recognizing workload simply transfer the burden to committed individuals.

3.3 Patchwork, Portfolios, and Programmatic Assessment: Dialogue Across Time

The cases examined so far have focused largely on assessment dialogue within or around particular tasks: a draft essay, a co-created rubric, a feedback tutorial, a peer review workshop, a negotiated assignment. These practices matter because they interrupt the one-way flow of conventional assessment. Yet dialogic assessment cannot reach its full educational potential if it remains confined to isolated episodes. Students do not develop disciplinary judgement, professional identity or self-regulation through one feedback conversation. They develop through repeated encounters with standards, evidence, feedback, reflection and revision over time. The temporal dimension of dialogic assessment is therefore crucial.

This section examines assessment designs that extend dialogue across a programme rather than treating each assignment as a separate measurement event. Portfolios, patchwork texts, programmatic

assessment and ipsative feedback all shift attention from single performances to trajectories of development. They ask not only “How good is this piece of work?” but “What pattern of growth does this body of evidence reveal?” “How has the student responded to feedback?” “What judgement can be made across time?” and “How is the learner coming to understand their own development?” These designs are dialogic not simply because they involve conversation, but because they create a continuing relationship among students, mentor, evidence and standards.

1. Portfolios and Patchwork Texts: Dialogue With an Evolving Body of Evidence

Portfolio assessment is one of the most established longitudinal assessment forms. At its simplest, a portfolio is a purposeful collection of evidence selected to demonstrate learning, development or competence. Yet the educational value of the portfolio does not lie merely in accumulation. A folder of disconnected tasks is not a portfolio in the stronger sense. Portfolio assessment becomes dialogic when the student selects evidence, interprets its significance, reflects on development, receives mentor response and revises the account of learning over time.

Klenowski’s (2002) work on portfolio assessment is important because it frames portfolios not only as repositories but as learning-oriented assessment practices. Portfolios allow students to gather evidence of achievement, make choices about representation, and construct an account of learning. They require students to ask: What counts as evidence? What does this piece of work show? How does it relate to criteria? What has changed since earlier work? In this way, the portfolio shifts assessment from isolated performance to narrated development.

Reflective components are crucial. Without reflection, portfolios risk becoming bureaucratic collections: assignments, certificates, feedback forms, observation reports, artefacts and checklists assembled for inspection. Reflection turns evidence into argument. The student must

explain why particular evidence matters, how it demonstrates learning, what feedback has been used, what remains unresolved and how development has occurred. The assessor or mentor then responds not only to the artefacts but to the student's interpretation of them.

Baume and Yorke's (2002) study highlights the need to examine the reliability of portfolio judgements. Reflective integration remains an important design aim. The portfolio can support learning when students are required to make connections among experiences, tasks and standards. It is not enough to show that several pieces of work have been completed. The student must construct coherence across them. In a programme where modules are otherwise fragmented, the portfolio can become a site where students make sense of the curriculum as a developmental whole.

Patchwork text assessment, associated with Scoggins and Winter (1999), offers a related but distinctive model. In the patchwork text, students produce a series of short "patches" over time: brief pieces of writing, reflections, analyses, field observations, responses to readings or practice-based accounts. These patches are not treated as isolated submissions. They are later "stitched" together through a final integrative commentary. The metaphor is important: the educational work lies not only in producing each patch but in constructing relationships among them.

The patchwork text is particularly suited to dialogic assessment because it creates repeated occasions for feedback and reflection. A student may write an early patch on a theoretical concept, a later patch applying that concept to practice, another reflecting on a seminar debate and a final commentary explaining how their understanding has changed. The tutor's role is not merely to mark each fragment, but to help the student see emerging patterns. Dialogue occurs between student and teacher, but also between earlier and later versions of the student's own thinking.

This makes the patchwork text a temporal pedagogy. It resists the fiction that learning appears fully formed in a final essay. Instead, it allows learning to be seen as cumulative, recursive and sometimes uneven.

Students may revise earlier assumptions, recognize contradictions, return to feedback, or reinterpret previous work from a later standpoint. The final integrative commentary becomes an act of evaluative judgement: the student must decide what their own body of work means.

Portfolios and patchwork texts therefore reconfigure assessment in at least three ways. First, they expand the object of judgement from a single product to an evolving body of evidence. Second, they require students to participate in the interpretation of that evidence. Third, they create space for mentoring dialogue across time. Their weakness, however, is equally clear: without careful design, they can become paperwork-heavy, reflective in name only, or dependent on uneven mentor support.

2. Programmatic Assessment: From Single-Shot Measurement to Longitudinal Decision-Making

Programmatic assessment, developed most extensively in medical and health professions education by van der Vleuten, Schuwirth and colleagues (2012), represents one of the most systematic alternatives to single-shot measurement. Its core principle is that no single assessment should carry disproportionate weight in judging complex competence. Instead, programmes gather multiple low-stakes data points over time, each providing feedback and evidence. High-stakes decisions are then made holistically by aggregating this evidence across contexts, assessors and methods.

This model directly challenges the measurement logic of traditional assessment. In the conventional model, a major examination or final task often functions as the decisive evidence of achievement. Programmatic assessment argues that complex competence cannot be validly inferred from isolated performances. A student may perform well on one exam but poorly in practice; another may struggle early but show sustained improvement. Competence is better understood as a pattern across time.

In programmatic assessment, individual assessment moments are primarily learning opportunities. A workplace observation, short test, case

discussion, reflective entry, simulation, peer review, supervisor report or self-assessment may all count as data points. Each data point is intended to carry limited decision-making weight on its own while providing useful feedback. The student is expected to review the evidence, identify learning needs, discuss progress with a mentor and set goals. The high-stakes decision is postponed until sufficient evidence has accumulated.

This structure is dialogic in several senses. First, feedback is distributed across time rather than concentrated at the end. Second, students must engage with evidence repeatedly rather than passively receive grades. Third, mentors or coaches support interpretation. Fourth, final judgement is holistic and committee-based rather than mechanically derived from one score. The programme becomes a continuing conversation about development.

The medical education context is important because it reveals why such a model is needed. Professional competence includes knowledge, technical skill, communication, ethical judgement, teamwork, clinical reasoning and responsiveness to feedback. These cannot be captured adequately by a single written exam or isolated performance. Programmatic assessment therefore aligns assessment with the complexity of professional practice.

The model aims to separate individual feedback encounters from immediate pass/fail decisions. Low-stakes data points are intended to support learning, with progression decisions based on a broader pattern of evidence. This separation is not automatic: Schut et al. (2018) found that learners could still perceive intended low-stakes assessments as high stakes. Clear decision rules, opportunities to improve and supportive relationships are therefore essential.

However, programmatic assessment does not remove judgement. It relocates judgement from single events to accumulated evidence. High-stakes decisions still occur, but they are made through holistic review. This requires robust procedures: clear outcomes, diverse evidence, trained assessors, mentoring structures, portfolio systems and decision

committees. The legitimacy of the model depends not on abandoning standards, but on making longitudinal judgement trustworthy.

3. Ipsative and Longitudinal Feedback: Judging Progress Against the Self

Ipsative assessment adds another dimension to longitudinal dialogue. Whereas normative assessment compares students with one another, and criterion-referenced assessment compares performance with external standards, ipsative assessment compares a student's current performance with their own prior performance. The central question becomes: How has this student developed?

Hughes's (2014) work on ipsative assessment is important because it foregrounds progress as an assessment focus. In many higher education systems, students receive marks on separate tasks but do not receive systematic feedback on their trajectory. They may know that they scored 58 on one essay and 64 on another, but not what kind of development this represents or how their learning profile is changing. Ipsative feedback creates a developmental conversation by connecting current work to previous work.

This does not mean abandoning standards. A student's improvement from weak to slightly less weak work cannot automatically be treated as high achievement. External standards remain necessary. But ipsative assessment adds a dimension often missing from conventional grading: recognition of movement. It asks students to attend to growth, persistence, strategy change and responsiveness to feedback. It can therefore be motivating, especially for students who are improving but remain below the highest bands.

In dialogic terms, ipsative feedback sustains a conversation between the student's past, present and possible future performance. A tutor might say: "In your first essay, evidence was mostly descriptive; in this assignment, you are beginning to use evidence to support claims, but the link to theory remains underdeveloped." Such feedback does more than

judge the current task. It gives the student a narrative of development. It shows what has changed and what the next developmental step might be.

Ipsative approaches are especially powerful when linked to portfolios or programme-level advising. Students can review feedback across tasks, identify recurring patterns and set goals. For example, a student in a social science degree may track development in argumentation, source integration, methodological reasoning and conceptual clarity across several modules. The assessment conversation then becomes cumulative. The student learns to see development as an object of reflection.

This longitudinal self-referenced perspective also supports self-regulation. Students who understand their own learning trajectory are better positioned to plan, monitor and evaluate future work. They can identify persistent weaknesses, recognize progress and seek targeted feedback. Ipsative assessment therefore contributes to the development of evaluative judgement by asking students to judge not only products but growth.

4. The Mentor, Tutor or Academic Advisor as Dialogic Partner

Longitudinal assessment requires a role that single-task assessment often lacks: the mentor, tutor, coach or academic advisor who helps the student interpret evidence over time. In portfolio and programmatic assessment, this role is central. Without mentoring, portfolios can become bureaucratic archives. With mentoring, they become learner charts: records of development that support reflection, feedback and decision-making.

The mentor's work is interpretive. Students may possess multiple feedback comments, grades, observation reports, peer reviews and reflective entries, but still be unable to make sense of them. The mentor helps identify patterns. Are weaknesses recurring? Has feedback been acted upon? Is progress visible? Are there contradictions between self-perception and external evidence? What goals should be set? The mentor

does not simply tell the student what to do; the mentor supports the student in constructing a coherent account of development.

This role requires data infrastructure. Longitudinal dialogue depends on access to evidence across time. If feedback is scattered across separate modules, platforms and teachers, neither student nor mentor can easily construct a developmental narrative. Portfolios, learning management systems, advising records and assessment dashboards can support this work if they are designed for interpretation rather than mere storage. The question is not simply whether data exists, but whether it can be used meaningfully in conversation.

The mentor role also requires relational infrastructure. Students are more likely to engage honestly with weakness if they trust the person helping them interpret evidence. This is particularly important in contexts where the mentor also has an assessment role. The combination of support and judgement can be productive, but it can also create tension. Students may conceal difficulties if they believe disclosure will harm progression. Mentors may struggle to balance advocacy with accountability.

Loosveld et al. (2020), drawing on interviews with 18 faculty mentors, identified four mentoring positions: facilitator, coach, monitor and exemplar. These describe different beliefs about mentoring, not a ranking of effectiveness. In their account, the monitor is a collaborative critical friend, while the exemplar is more directive. The practical implication is to make mentoring roles explicit and support mentors in balancing care, challenge and developmental responsibility.

Equity is also crucial. The quality of mentoring can vary considerably. Some students receive rich guidance; others experience perfunctory meetings. Some mentors are skilled in feedback dialogue; others focus narrowly on documentation. If longitudinal assessment depends heavily on mentoring, then inequity in mentoring quality becomes an assessment equity issue. Programmes must therefore train mentors, monitor

mentoring quality and ensure that students have access to support if relationships fail.

5. Mechanisms of Longitudinal Dialogic Assessment

Longitudinal assessment designs work through several mechanisms. The first is continuity. Dialogue becomes more powerful when it is not reset after every task. A mentor who has seen earlier work can interpret current performance developmentally. A student who revisits earlier feedback can recognize patterns. A programme that connects tasks can support cumulative learning. Continuity allows feedback to become a thread rather than a series of disconnected comments.

The second mechanism is narrative coherence. Students often experience higher education as a sequence of modules, deadlines and grades. Longitudinal assessment invites them to construct a story of development. This story is not a fiction imposed on learning; it is an interpretive act that helps students connect experiences. The portfolio commentary, patchwork synthesis or mentoring meeting asks: What has changed? What evidence supports that claim? What remains difficult? What kind of learner or professional are you becoming?

The third mechanism is the decoupling of every-task-is-graded pressure. When each assessment is high stakes, students are incentivized to avoid risk, conceal weakness and focus on marks. Low-stakes data points can create a different assessment culture. Students can receive feedback on emerging competence, experiment with strategies, and use evidence formatively. High-stakes judgement is not abolished, but it is delayed until a richer picture has developed.

The fourth mechanism is the cultivation of self-regulation. Longitudinal designs require students to review evidence, set goals, monitor progress and act on feedback. These are core self-regulatory processes. Over time, students ideally become less dependent on external prompts and more capable of managing their own development. The mentor gradually shifts responsibility back to the learner.

The fifth mechanism is triangulation. Complex learning is better understood through multiple forms of evidence. A written exam, practical observation, reflective commentary, peer feedback and supervisor report each reveal different aspects of performance. Dialogue across evidence helps students and assessors see strengths and weaknesses more holistically. This is especially important in professional and interdisciplinary programmes.

3.4 Peer Dialogue and Calibration: Assessment as Distributed Conversation

If dialogic assessment is to become more than an ideal for small seminars and intensive tutorials, it must find ways to distribute assessment conversation beyond the teacher–student dyad. Peer-based dialogue is therefore central to any scalable account of dialogic assessment. It offers a way of multiplying opportunities for students to encounter standards, test judgements, compare work, articulate reasons and receive response without making the teacher the sole source of evaluative information. Properly designed, peer review, peer feedback, peer assessment, comparative judgement and group assessment can transform students from recipients of judgement into participants in a community of judgement.

The case for peer dialogue should not be framed merely as a workload solution. It is true that peer processes can reduce exclusive dependence on teacher comments, particularly in large cohorts. But their deeper value is developmental. Students learn standards by applying them. They learn feedback by producing it. They learn judgement by comparing, explaining, disagreeing and revising. Peer-based assessment dialogue therefore addresses one of the central problems identified throughout this book: standards cannot be transferred by documents alone. They must be practised socially.

At the same time, peer dialogue is fragile. It can become superficial, inequitable, resisted or unreliable. Students may distrust peers, resent being asked to do what they see as the teacher's job, or avoid honest critique to preserve relationships. Participation can be uneven. Cultural, linguistic and confidence differences can shape who speaks, whose judgements are taken seriously and whose work receives useful response. Peer dialogue is therefore not automatically democratic or educational. It must be designed.

1. Peer Review and Peer Feedback as Dialogue

Peer review and peer feedback are among the most widely used forms of peer-based assessment dialogue. In their simplest form, students read or observe one another's work and offer comments for improvement. Yet the educational value of peer review does not lie only in the feedback received. Nicol, Thomson and Breslin's (2014) study of students' accounts suggests that producing peer feedback can support evaluative judgement and reflection on one's own work. It does not establish a universal ranking of the benefits of giving and receiving feedback. This claim is important because it shifts attention from peer feedback as a substitute for teacher feedback to peer review as a practice of judgement.

When students review another's work, they must interpret the task, apply criteria, notice features of quality, identify weaknesses, formulate reasons and communicate suggestions. These activities require evaluative judgement. A student who comments that a peer's paragraph "needs more evidence" must implicitly decide what claim is being made, what kind of evidence would count, whether the existing evidence is sufficient and how the paragraph might be strengthened. Even if the comment is imperfect, the act of making it can sharpen the reviewer's understanding of their own work.

This explains why giving feedback can generate internal feedback. While examining a peer's draft, students often notice problems that are also present in their own writing. They see another student's unclear

introduction and return to their own introduction with new questions. They recognize that a peer has integrated theory more effectively and begin to understand what their own essay lacks. The comparison between peer work and one's own work becomes a feedback-generating mechanism.

Peer feedback also makes standards social. In teacher-only assessment, students may experience standards as private properties of the assessor. In peer review, standards become topics of shared discussion. Students ask: Is this argument clear? Does this evidence support the claim? Is the method justified? Is the reflection sufficiently critical? Such questions make quality visible as something that can be reasoned about rather than simply announced by authority.

However, peer review is dialogic only when students can interact with one another's interpretations. A one-way exchange of written peer comments may be useful, but it can still replicate transmission. Dialogue requires opportunities for clarification, questioning, comparison and response. The author may ask what a comment means; the reviewer may explain; other peers may disagree; the group may return to the criteria. Through this process, students learn that feedback is not simply information but negotiated meaning.

2. Calibration and Comparative Judgement

Peer-based dialogue also supports calibration: the process by which students' judgements become increasingly aligned with disciplinary standards and with the judgements of more experienced assessors. Calibration does not mean that students become identical to teachers or that all disagreement disappears. It means that students become better able to recognize quality, justify evaluations and understand where their judgements sit in relation to a community of practice.

Adaptive comparative judgement, associated with Pollitt (2012), offers a particularly important model. Rather than asking assessors to apply a rubric to a single piece of work, comparative judgement asks them to

compare two pieces and decide which is better. Repeated comparisons by multiple judges generate a rank order. The underlying insight is that people often find it easier to make comparative judgements than absolute ratings. It may be difficult to decide whether an essay is exactly 68 or 72, but easier to decide which of two essays offers the stronger argument.

The relevance to dialogic assessment is significant. Comparative judgement challenges the assumption that standards must always be expressed through detailed criteria before judgement can occur. Standards can also be developed through repeated comparison. Students learn quality by seeing differences. They encounter pairs of work, make judgements, compare their reasoning with others and gradually refine their sense of what matters. This is not rubric-free relativism. It is a different route into standards: one grounded in comparative perception and collective judgement.

Adaptive comparative judgement can produce consistent rankings under suitable conditions, but reliability alone does not establish validity. Mentzer, Lee and Bartholomew (2021), for example, found that peer comparative judgements in a design-thinking course were not significantly correlated with instructors' rubric grades for the same task. Comparative judgement therefore requires evidence that judges attend to the intended qualities; consistency cannot substitute for this check.

Calibrated peer review works on related principles. Students may first review benchmark examples, compare their judgements with instructor judgements, and receive feedback on the quality of their evaluations before reviewing peers. Calibration makes peer review more trustworthy because it reduces the risk that students apply idiosyncratic or mistaken standards. It also makes the learning purpose explicit: students are not merely helping classmates; they are learning to judge.

These approaches are especially valuable for complex performances where holistic judgement matters: essays, designs, portfolios, presentations, reflective accounts, project proposals and creative work. Rubrics can help orient attention, but comparison reveals how criteria

interact. A student may see that one essay is not stronger because it has more sources, but because it uses sources to develop a more coherent argument. Such insight is difficult to acquire from descriptors alone.

3. Designing Productive Peer Dialogue

Peer dialogue must be designed if it is to become genuinely deliberate rather than perfunctory. Simply placing students in groups and asking them to “give feedback” is rarely enough. Students need training, protocols, norms and repeated practice.

Training should address both the cognitive and interpersonal dimensions of peer feedback. Students need to know what quality looks like, how to use criteria, how to distinguish description from evaluation, how to make suggestions specific, and how to justify comments with evidence from the work. They also need to learn how to give critique respectfully, how to receive critique without defensiveness, and how to ask clarifying questions. Feedback literacy is therefore not only a student’s ability to use teacher comments; it includes the ability to participate responsibly in peer evaluative dialogue.

Anonymity choices matter. Anonymous peer review can reduce fear of offending classmates and may encourage more honest critique. It can also reduce social bias in some contexts. However, anonymity can weaken dialogue because students cannot easily ask follow-up questions or build trust. Named peer review can support accountability and richer conversation, but may lead to politeness, reluctance or interpersonal tension. The best choice depends on task, cohort, culture and purpose. Early peer review may benefit from anonymity or semi-anonymity; later developmental dialogue may require named interaction.

Structured protocols are essential. A useful protocol might ask reviewers first to summarize the author’s main claim, then identify one strength, then ask one genuine question, then suggest one priority for revision. Another protocol might require comments to be linked to specific criteria and evidence in the text. Protocols reduce vague praise and

unsupported criticism. They also protect students from being overwhelmed by too many disconnected comments.

Mercer's account of exploratory talk and the work on interthinking developed with Littleton (Littleton & Mercer, 2013) provide a strong theoretical basis for productive peer dialogue. Mercer distinguishes exploratory talk from disputational talk, where participants compete and disagree unproductively, and cumulative talk, where participants build agreement uncritically. Exploratory talk is characterized by critical but constructive engagement: reasons are offered, claims are challenged, alternatives are considered, and participants work toward shared understanding. This is precisely the kind of talk peer assessment requires.

In peer feedback, exploratory talk might look like students debating whether a paragraph is descriptive or analytical, using evidence from the text to support their views. One student may argue that the paragraph is analytical because it compares two authors; another may respond that comparison alone is not analysis unless it advances the essay's argument. Through such exchange, the group does more than evaluate one draft. It develops a shared understanding of analysis.

Interthinking names the collective cognitive work that occurs through such talk. Students think together. They use language to make reasoning public, test interpretations and build knowledge that no individual may have produced alone. Peer dialogue becomes powerful when it creates this intermental space. But exploratory talk does not occur automatically. Students must be taught ground rules: listen actively, ask for reasons, challenge ideas not people, invite quieter voices, use evidence, seek shared understanding and remain open to revision.

4. Group and Team Assessment as Dialogic Sites

Group and team assessment introduces another form of peer-based dialogue. Here students do not only comment on one another's work; they negotiate the production of shared work and often evaluate contribution. Group projects, presentations, design tasks, laboratory reports, fieldwork projects and professional simulations all create assessment situations in which students must coordinate roles, standards and responsibility.

The dialogic potential of group assessment lies in negotiation. Students must decide what the task requires, what quality will look like, who will do what, how contributions will be integrated and how disagreements will be resolved. These negotiations can develop metacognition and judgement because students must articulate expectations. A team designing a research presentation must ask: What is our central argument? What evidence is most persuasive? How do we divide speaking roles fairly? What standard will we use to judge whether the presentation is coherent?

Peer accountability is central. Group assessment is often criticized because of free riding: some students contribute less while receiving the same mark. Peer assessment of contribution can address this problem by making participation visible. Students may evaluate one another's reliability, quality of contribution, collaboration, communication or leadership. Such systems can encourage responsibility, but they also introduce risks. Students may inflate ratings to avoid conflict, punish disliked peers, or misunderstand invisible labour. Peer accountability must therefore be carefully structured, with clear criteria and opportunities for reflection.

Group assessment also causes conflict. Conflict is not always a failure; it can be a site of learning. Students must negotiate standards, resolve disagreement and manage difference. However, conflict becomes damaging when unsupported. Teachers need to provide mechanisms for early reporting, role negotiation, team contracts, interim check-ins and

reflective accounts. Without such structures, group assessment can reproduce inequities, with confident students dominating and quieter students marginalized.

Team assessment is especially relevant to professional education because many professions require collaborative judgement. Healthcare, engineering, education, social work, business and design all involve working with others under constraints. In such contexts, peer dialogue is not only a pedagogical strategy but part of the competence being assessed. The ability to give, receive and negotiate feedback is itself a professional capability.

5. Mechanisms and Evidence

Peer dialogue develops judgement through several mechanisms. First, it increases exposure to variation. Students see multiple ways of responding to the same task. This variation is educationally valuable because it helps them understand that quality is not a single template. They learn by comparing stronger and weaker approaches.

Second, peer dialogue requires articulation. To give feedback, students must turn tacit impressions into language. They must explain why something works or does not work. This articulation strengthens understanding because vague preference must become reasoned judgement.

Third, peer dialogue creates reciprocal perspective-taking. Students experience the role of assessor and are assessed. As assessors, they learn the difficulty of judging fairly. As authors, they learn how comments are received and interpreted. This dual perspective can increase feedback literacy and empathy.

Fourth, peer dialogue supports internalization. Through repeated peer review, students begin to carry evaluative questions into their own work. They ask themselves the kinds of questions they have asked others. This is the developmental movement from external dialogue to internal judgement.

The evidence on agreement between peer and teacher assessment is cautiously supportive but context-dependent. Li et al.'s (2016) meta-analysis found a moderately strong average correlation between peer and teacher ratings. Correlation, however, does not mean exact agreement in marks, nor does it by itself establish the validity of either judgement. Training, clear criteria and multiple perspectives are useful safeguards, and high-stakes peer assessment needs local checks for bias, consistency and alignment with the intended learning outcomes.

Student trust in peers remains a major issue. Many students initially believe that assessment is the teacher's job because teachers possess expertise and formal authority. This resistance is understandable. Students may fear receiving poor advice, being judged unfairly or having their grade affected by peers who lack competence. To address this, teachers must explain the purpose of peer assessment. If peer review is framed as replacing teacher judgement, students may resist. If it is framed as developing judgement and generating feedback for revision, students are more likely to engage.

Trust also grows through calibration. When students see that their judgements can become more aligned with teacher judgements, confidence increases. When they experience useful feedback from peers, resistance declines. The legitimacy of peer dialogue must therefore be earned through design, not asserted rhetorically.



Mechanisms and Dynamics

A Theoretical Framework for Dialogic Assessment

4.1 Trust, Power, and Vulnerability: The Relational Engine of Assessment Dialogue

The preceding chapters have examined dialogic assessment as design: feedback loops, co-created criteria, peer review, portfolios, programmatic assessment and longitudinal mentoring. Yet none of these practices works merely because it is present in a curriculum. A feedback tutorial can become a defensive exchange. Peer review can become polite performance. A portfolio meeting can become bureaucratic compliance. A negotiated rubric can become symbolic participation. The success or failure of dialogic assessment depends on mechanisms beneath the technique: trust, vulnerability, emotion, power, recognition and answerability.

This section argues that dialogic assessment is not only a cognitive process of making standards clear. It is a relational and affective practice through which students decide whether it is safe to disclose misunderstanding, whether critique can be trusted, whether they are recognized as capable knowers, and whether assessment authority is being exercised accountably. Dialogue produces learning only when these relational conditions are strong enough to support honest participation. Where they are absent, the same techniques may produce compliance, resistance, silence or strategic performance.

1. Trust as Precondition and Product

Trust is both a precondition for dialogic assessment and one of its possible outcomes. Students are more likely to engage with feedback when they believe that the assessor is competent, fair, invested in their development and willing to respond to them as more than a grade-producing object. Without trust, feedback is filtered through suspicion. A critical comment may be interpreted as hostility, bias, indifference or confirmation that the student does not belong. With trust, the same comment can be heard as a demanding but developmental invitation.

Carless's (2009) work on trust and assessment reform is important because it shifts attention from feedback text to feedback relation. Students do not evaluate comments in isolation. They ask, often implicitly: Does this teacher understand what I was trying to do? Is this judgement fair? Does this person want me to improve? Can I ask for clarification without looking foolish? Will my uncertainty be used against me? These questions shape whether feedback is read, believed and acted upon.

Trust enables disclosure. In dialogic assessment, students must often reveal what they do not understand: confusion about criteria, uncertainty about disciplinary expectations, difficulty interpreting a comment, or anxiety about performance. Such disclosure is educationally valuable because it gives teachers access to the learner's interpretive process. Yet students disclose only when the relationship feels safe enough. A student who fears humiliation may nod silently while misunderstanding persists. A student who fears penalty may conceal difficulty until final submission. A student who distrusts the assessor may treat feedback as something to defend against rather than think with.

High-stakes grading can discourage candour. When every assessment encounter is also a potential judgement with serious consequences, students become strategic. They may avoid asking questions that expose weakness. They may perform confidence rather than seek help. They may treat feedback dialogue as an opportunity to discover what the assessor

wants rather than as a space for inquiry. This does not mean students are manipulative; it means the assessment environment teaches them that vulnerability is risky. Dialogic assessment therefore requires careful handling of stakes. Low-stakes feedback points, draft conversations, formative peer dialogue and mentoring meetings can create spaces where students can speak more honestly because the immediate threat of grade penalty is reduced.

Trust is also produced through dialogic practice. When students experience feedback as responsive, specific and usable, trust grows. When teachers explain standards, acknowledge ambiguity, invite questions and follow through on commitments, students learn that dialogue is worthwhile. When peer review is well facilitated and produces helpful insights, students begin to trust peers as sources of judgement. Trust is therefore not a static personality trait or general classroom atmosphere. It is built through repeated assessment interactions in which critique proves fair, intelligible and actionable.

2. Vulnerability and Emotion in Assessment

Assessment is emotionally charged because it involves judgement. Students submit work that represents time, effort, identity, aspiration and often fear. To be assessed is to become visible. The student's thinking, writing, performance or professional conduct is opened to scrutiny. This exposure produces vulnerability, especially when the judgement is public, high-stakes, unexpected or linked to identity.

Feedback research increasingly recognizes that students do not process feedback as neutral information. They receive it affectively. A grade may produce relief, shame, anger, disappointment or pride before any comment is read. Negative emotions can block uptake. Shame may lead students to avoid feedback altogether. Defensiveness may lead them to reject critique. Anxiety may narrow attention to the mark. Anger may redirect focus from learning to perceived unfairness. Even accurate

feedback can fail educationally if the emotional conditions of reception prevent sense-making.

Pitt and Norton's (2017) interview study of 14 final-year undergraduates suggests that feedback engagement is shaped by how students feel about the judgement and about themselves as learners. Students may experience comments as personal even when teachers intend them as comments on work. This is especially true in disciplines where work expresses voice, interpretation or professional identity. A student receiving criticism on an essay, reflective account, creative project or teaching performance may hear not only "this piece needs improvement" but "you are not good at this" or "you do not belong here."

Molloy and Bearman (2019) examine the tension between showing vulnerability and maintaining credibility in health professions education. Their account of intellectual candour proposes that educators can make their own uncertainty discussable and invite reciprocal openness. Applied to assessment dialogue, this suggests a role for carefully framed conversations in which students can acknowledge difficulty without treating it as a final verdict on their capability.

Relational framing helps manage emotion. This does not mean softening all critique or avoiding difficult standards. Dialogic assessment requires challenge. But challenge must be framed as part of development rather than as final condemnation. Effective relational framing might include making the purpose of feedback explicit, distinguishing the work from the person, identifying strengths as resources for improvement, inviting the student's interpretation, and connecting critique to future action. The question is not how to remove emotion from assessment but how to create conditions in which emotion does not prevent learning.

Teachers also experience emotion in assessment. They may feel frustration when students ignore feedback, anxiety about fairness, defensiveness when challenged, exhaustion from marking, or discomfort in giving critical comments. Dialogic assessment requires emotional

labour from educators. They must hold the tension between care and judgement, empathy and standards, responsiveness and authority. Institutions often underestimate this labour, treating feedback as a technical task rather than a demanding relational practice.

3. Power, Asymmetry and the Examination

A relational account of assessment must also be a power-conscious account. Dialogue does not occur between equal parties. Teachers design tasks, interpret standards and award grades. Institutions classify students, determine progression and confer credentials. Students may ask questions, negotiate meanings and participate in feedback, but they do so within a structure in which others hold formal authority over their academic futures.

Foucault's analysis of the examination in *Discipline and Punish* (1977) remains powerful for understanding this structure. The examination combines hierarchical observation and normalizing judgement. It makes individuals visible, records them, compares them and produces knowledge about them. In education, assessment does not merely measure learning; it creates categories of student: excellent, failing, borderline, at risk, competent, deficient, progressing. These categories are not neutral labels. They shape how students are seen by institutions and how students come to see themselves.

The examination is therefore a disciplinary technology. It normalizes by defining expected performance and identifying deviations. It surveils by requiring students to submit work for scrutiny. It individualizes by producing records, marks, rankings and profiles. It disciplines by encouraging students to regulate themselves according to assessment expectations. This is why assessment can produce compliance as well as learning. Students may learn to write for the rubric, perform for the examiner, conceal uncertainty and avoid intellectual risk because the assessment regime rewards conformity.

A Foucauldian reading does not imply that all assessment is oppressive or that standards should be abandoned. Rather, it reveals that assessment is never innocent. It operates through power. The key question for dialogic assessment is therefore not whether power can be removed, but how it can be made visible, accountable and educative.

Dialogue makes power visible by naming roles and stakes. The teacher does not pretend that the relationship is equal. The student knows who will mark the work, what standards apply and what can or cannot be negotiated. Dialogue makes power accountable by requiring reasons. The assessor must explain judgement, respond to questions, clarify standards and acknowledge uncertainty where it exists. Dialogue makes power educative by using the asymmetry between expert and novice as a resource for development. The teacher's authority becomes a means of inducting students into judgement rather than simply classifying them.

This is the difference between authoritarian and authoritative assessment. Authoritarian assessment demands acceptance because the assessor has power. Authoritative dialogic assessment invites understanding because the assessor has responsibility. It retains standards, but it opens their meaning to inquiry. It retains judgement, but it makes judgement answerable.

4. Recognition, Identity and Belonging

Assessment is also an act of recognition. When teachers respond to student work, they do more than identify quality. They recognize, or fail to recognize, the student as a developing knower. This recognition has consequences for identity and belonging. Students learn not only what grade they received, but what kind of participant they are becoming in an academic community.

Honneth's (1995) theory of recognition offers a useful ethical framework. Human identity is formed through relations of recognition: being cared for, respected and socially esteemed. Misrecognition damages self-relation. In assessment contexts, misrecognition can occur when

feedback treats students as deficient objects, when criteria ignore their background knowledge, when comments dismiss rather than develop their contribution, or when institutional processes reduce them to marks and risk categories.

Recognition does not mean praise. A student can be recognized through demanding critique if the critique acknowledges their capacity to grow. Conversely, a student can be misrecognized through empty positivity if the feedback avoids serious engagement with their work. Recognition means that the student is addressed as someone whose thinking matters, whose effort is interpretable, whose misunderstandings are developmental, and whose participation in the discipline is possible.

This is particularly important for students whose belonging in higher education is fragile. First-generation students, minoritized students, international students, disabled students and students returning after failure may read assessment encounters as signals of inclusion or exclusion. A dismissive comment may confirm that they are outsiders. A dialogic conversation that takes their work seriously may support a sense of academic legitimacy. Assessment therefore participates in the formation of academic identity.

Bakhtin's (1993) concept of answerability complements recognition theory. Every evaluative utterance is addressed to someone and carries responsibility for how it enters the other's world. A feedback comment is not merely an informational statement; it is an answer to a student's work and an invitation to further response. The assessor is answerable for the effects of this utterance, not in the sense that teachers can control every reaction, but in the sense that assessment speech is ethically consequential.

Answerability requires specificity. The assessor responds to this student, this work, this attempt, this moment of development. Generic feedback can fail ethically because it does not answer the student's actual performance. Overly standardized comments may satisfy administrative efficiency but leave the student unrecognized. Dialogic feedback, by

contrast, seeks to answer the students' questions in a way that opens further action.

Recognition and answerability together frame assessment as an ethical relationship. The student is not merely being measured; the student is being addressed. The assessor is not merely applying criteria; the assessor is responding to a developing person through the medium of academic work. This does not remove standards. It deepens the responsibility with which standards are enacted.

5. A Model of Relational Conditions

The relational and affective mechanisms discussed above can be synthesized into a model of the conditions under which dialogic assessment produces learning rather than compliance or resistance. Four conditions are central: trust, safety, continuity and recognition.

Trust concerns the learner's belief that assessment dialogue is fair, credible and oriented toward development. Without trust, critique is discounted or defended against. With trust, students are more likely to risk engagement with difficult feedback.

Safety concerns the student's capacity to disclose uncertainty without excessive threat. Psychological safety does not mean absence of challenge; it means that challenge occurs in a context where mistakes can be discussed and used. Without safety, students perform competence. With safety, they can reveal the misunderstandings assessment needs to address.

Continuity concerns the persistence of relationships and evidence over time. One-off feedback encounters may be useful, but trust and developmental insight are strengthened through repeated interaction. Continuity allows the teacher or mentor to see growth and allows the student to experience feedback as part of a trajectory rather than a verdict.

Recognition concerns being addressed as a capable participant in learning. It involves respect for the student's agency, seriousness toward their work, and acknowledgement of their developing identity. Without

recognition, assessment becomes objectifying. With recognition, judgement can support belonging and self-authorship.

These conditions interact. Trust supports safety; safety enables candour; candour makes better feedback possible; better feedback builds trust. Continuity strengthens recognition because the student is seen over time rather than through a single performance. Recognition makes critique more tolerable because students feel judged as developing learners rather than dismissed as inadequate.

Where these conditions are absent, dialogic assessment techniques may fail. A feedback conference without trust may become defensive. Peer review without safety may become polite and useless. Co-created criteria without recognition may privilege confident voices. Programmatic mentoring without continuity may become bureaucratic monitoring. Dialogue cannot be reduced to interactional format. It depends on the quality of the relation in which interaction occurs.

4.2 From Telling to Internalising: How Dialogue Builds Evaluative Judgement and Self-Regulation

The previous section argued that dialogic assessment works or fails through relational and affective conditions: trust, safety, recognition, continuity and accountable power. This section turns to the cognitive and developmental mechanisms by which dialogue produces durable learning. If Section 4.1 explained why students become willing to engage with assessment dialogue, this section explains how such engagement becomes internal capacity. The central claim is that dialogic assessment matters because it enables students to internalize standards, generate internal feedback, regulate their own learning and eventually judge quality without continuous teacher presence.

The argument is developmental. Dialogic assessment is not merely a better way of communicating teacher judgement. It is a mechanism for transforming external, socially mediated judgement into internal

evaluative capability. Students first encounter standards between people: in feedback conversations, peer review, exemplar discussions, moderation exercises, draft tutorials and mentoring meetings. Through repeated participation, they gradually appropriate the language, criteria, comparisons and habits of judgement. What begins as social dialogue becomes self-dialogue. What begins as teacher-supported judgement becomes independent evaluative judgement.

This developmental account connects sociocultural theory, self-regulated learning theory, internal feedback theory, calibration research and sustainable assessment. Taken together, these traditions show that dialogue is not only a moral or relational ideal. It is cognitive architecture for learning to judge.

1. The Internalization Mechanism

Vygotskian sociocultural theory offers a lens for explaining how dialogic assessment may support individual capability. Vygotsky's central proposition is that higher mental functions appear first on the social plane and only later on the psychological plane. They are first interpsychological, between people before becoming intrapsychological, within the individual. This book extends that principle to evaluative judgement as a theoretical interpretation.

Students are not born knowing how to judge academic quality. Nor do they acquire this capacity simply by reading criteria. They learn judgement by participating in the activity of judging with others. A teacher explains why an argument is underdeveloped; peers compare two drafts; a tutor asks how evidence supports a claim; a student revises after feedback; a group debates which exemplar better meets the standard. In these moments, judgement is external, social and mediated. The student is not yet fully independent but is participating in a practice of judgement that exceeds their current individual capacity.

Over time, this external dialogue can become internal dialogue. A student who repeatedly hears the question "What is the claim doing

here?” may begin to ask that question while drafting. A student who has compared several literature reviews may begin to notice when their own review is merely listing sources. A student who has participated in peer review may begin to anticipate how a reader will respond. The language of assessment becomes part of the student’s self-monitoring.

This is the internalization mechanism at the heart of dialogic assessment. Dialogue does not simply clarify standards; it provides the social form through which standards can be appropriated. The novice first borrows the evaluative language of the community, then uses it with guidance, then applies it to others’ work, then turns it toward their own work, and eventually uses it independently. Internalized judgement is therefore not private intuition. It is socially formed judgement that has become available for self-regulation.

Wertsch’s (1998) emphasis on mediated action helps specify this process. Internalization does not occur through direct transfer from teacher mind to student mind. It occurs through cultural tools: language, rubrics, exemplars, feedback comments, peer discussion, disciplinary concepts, comparison tasks and reflective prompts. These tools mediate action. A rubric does not create understanding by itself, but it can support understanding when used in dialogue. An exemplar does not teach automatically, but it becomes a mediational tool when students compare, discuss and apply it. Feedback does not internalize itself; students appropriate it through use.

This explains why dialogic assessment must be designed as repeated participation rather than one-off explanation. A single feedback conversation may help, but internalization requires cycles. Students need recurring opportunities to encounter standards in action, practise judgement, receive response and revise understanding. The developmental pathway is cumulative: from teacher-held standards to shared standards, to negotiated standards, to internalized judgement.

2. Dialogic Feedback and Self-Regulated Learning

Self-regulated learning theory offers a second mechanism. Zimmerman's (2002) model is especially useful because it describes learning as a cyclical process involving forethought, performance and self-reflection. In the forethought phase, learners analyse tasks, set goals and plan strategies. In the performance phase, they monitor progress and use strategies. In the self-reflection phase, they evaluate outcomes and make attributions that shape future learning. Dialogic assessment can feed each part of this cycle.

In the forethought phase, dialogue helps students understand what the task requires. Assignment briefings, exemplar discussions, criteria workshops and feedback request sheets help students form goals. Without such dialogue, students may begin with weak or mistaken task representations. They may think the task is asking for summary when it requires argument, or opinion when it requires evidence-based analysis. Dialogic clarification at the beginning of the cycle improves goal-setting.

In the performance phase, dialogue supports monitoring. Draft tutorials, peer review, self-assessment checklists and formative feedback allow students to compare work-in-progress against standards. The student asks: Am I doing what the task requires? Is the argument coherent? Is the method justified? Have I acted on previous feedback? Such questions are self-regulatory, but they are often learned through external prompts before becoming internal habits.

In the self-reflection phase, dialogue helps students interpret outcomes. A grade alone may produce emotion without understanding. Feedback dialogue can help students ask why the result occurred, what patterns are visible and what should change next time. Reflection becomes more productive when it is connected to evidence and future action. The cycle then begins again, with improved forethought.

Nicol and Macfarlane-Dick's feedback principles link feedback directly to self-regulation. Good feedback clarifies what good performance is, facilitates self-assessment, provides high-quality information, encourages

dialogue, supports motivation, provides opportunities to close the gap and informs teaching. These principles show that feedback is not merely corrective. Its deeper function is to help students become capable of regulating their own work.

Dialogic feedback does this by moving students from dependence on external judgement toward the generation of internal feedback. A student initially waits for the teacher to identify problems. Later, after repeated exposure to feedback dialogue, the student begins to identify likely problems before submission. Eventually, the student can monitor and adjust work without immediate teacher intervention. This is not the disappearance of feedback, but its transformation. Feedback becomes increasingly internal.

3. Comparison as the Core Cognitive Act

Nicol's recent work on internal feedback deepens this account by identifying comparison as the core cognitive act in learning from feedback. Students generate internal feedback when they compare their work, performance, understanding or strategies against some reference information. That reference information might be a teacher comment, an exemplar, a peer's draft, a rubric, a model answer, a previous attempt, a professional standard, or even a remembered classroom discussion.

This reframes feedback theory. Feedback is not only information given to students; it is information generated by students through comparison. A teacher comment can prompt comparison, but so can a peer text. An exemplar can prompt comparison. A dialogue can prompt comparison. The learner's mind produces internal feedback by noticing similarity, difference, gap, possibility or error.

For example, a student comparing their own essay introduction with an exemplar may notice that the exemplar does not merely announce a topic but constructs a problem. This comparison generates internal feedback: "My introduction is descriptive; I need to establish a problem and position my argument." The teacher has not written this comment.

The student has generated it through comparison. Similarly, a student listening to peer feedback may compare how several readers understood their argument and notice that the central claim is not as visible as intended.

Dialogue orchestrates richer comparisons. It does not leave students alone with reference information. It structures what they compare, how they compare, and what they notice. In an exemplar workshop, the teacher may ask students to compare two conclusions: Which one does more than repeat the essay? What evidence from the text supports your judgement? How does the stronger conclusion create significance? These questions guide comparison. Students learn not simply to look at examples, but to compare evaluatively.

Peer review also works through comparison. Students compare a peer's work with criteria, with their own work, with exemplars and with their understanding of the task. This produces internal feedback for both reviewers and receivers. The reviewer learns by judging; the receiver learns by seeing how others interpret the work. Comparative judgement methods make this mechanism explicit by asking students to compare pairs of performances and decide which better meets the standard. Repeated comparison develops a more refined sense of quality than abstract descriptors alone.

Comparison is therefore the cognitive bridge between external dialogue and internal judgement. Dialogue supplies reference points, language and prompts. Comparison generates internal feedback. Internal feedback supports self-monitoring and revision. Over time, this process becomes more automatic and sophisticated.

4. Calibration, Metacognition and the Novice Problem

A major challenge in assessment is novice miscalibration. Students who lack experience in a domain often cannot accurately judge the quality of their own work. They may overestimate because they do not yet know what they do not know, or underestimate because they cannot recognize emerging competence. This problem resembles what is often called the Dunning–Kruger effect: those with lower competence may also lack the evaluative resources needed to recognize low competence.

In higher education, miscalibration appears when students believe they have answered the question because they have written about the topic; when they assume that effort guarantees quality; when they interpret quantity of reading as depth of engagement; or when they believe a polished presentation compensates for weak analysis. Such misjudgements are not merely motivational problems. They are developmental. Novices lack internalized standards.

Dialogic assessment improves calibration through repeated judgement under guidance. Students compare their self-assessment with teacher assessment. They review peer work and then hear how the teacher interprets the same work. They apply rubrics to exemplars and discuss disagreements. They receive feedback on the quality of their feedback. Each cycle helps adjust their evaluative frame.

Self-assessment is central here, but only when treated as pedagogy rather than self-grading. Asking students to assign themselves a mark without training may produce unreliable judgements. Asking them to justify their judgement with reference to criteria, compare it with peer and teacher judgement, and revise their understanding can improve calibration. The developmental value lies in the reasoning process.

Peer assessment also supports calibration. Students encounter variation in quality and learn to distinguish levels of performance. When they compare their judgements with others, they discover where their standards are too lenient, too harsh or misdirected. Calibration is not

simply convergence with the teacher's mark. It is the growth of metacognitive awareness: students become more aware of how they judge and why.

Metacognition refers to thinking about one's own thinking. In assessment, metacognition includes awareness of goals, strategies, criteria, weaknesses, strengths and progress. Dialogic assessment supports metacognition by making judgement explicit. Students are asked not only to produce work but to explain how they judged it, what feedback they used, what changes they made and what remains uncertain. Reflection becomes an account of evaluative reasoning rather than a generic statement of effort.

Over time, this reduces novice miscalibration. Students become more accurate in identifying weaknesses, more specific in seeking feedback, and more strategic in improving work. They learn to ask better questions: not "Is this good?" but "Does my evidence actually support this claim?" not "What mark would this get?" but "Where is the argument least developed?" Such questions indicate growth in evaluative judgement.

5. Sustainable Assessment and the Long View

The purpose of internalization and self-regulation is not only better performance in the next assignment. It is sustainable learning. Boud's concept of sustainable assessment asks whether assessment equips students for learning beyond the immediate educational setting. The ultimate test of assessment is what students can do when no teacher is present to judge, correct or prompt them.

This long view is crucial for higher education. Graduates enter workplaces, professions, communities and research contexts where they must make judgements under uncertainty. They must decide whether work is good enough, whether evidence is credible, whether feedback from colleagues is useful, whether professional action meets a standard, or whether further learning is needed. If university assessment has trained

them only to comply with teacher requirements, it has failed as preparation for lifelong learning.

Sustainable assessment therefore shifts the purpose of feedback and judgement. The aim is not merely to improve the current product but to build the learner's capacity to judge future products. The aim is not merely to correct errors but to cultivate self-correction. The aim is not merely to certify learning but to develop the ability to continue learning.

Dialogue is the means to this end because independence is socially formed. Students do not become autonomous by being left alone. They become autonomous through supported participation in judgement, gradually taking over functions initially carried by teachers and peers. A mentor asks evaluative questions until the student can ask them. A peer review protocol models feedback until the student internalizes it. Exemplar discussions build comparison habits until the student can generate internal feedback independently.

Sustainable assessment also changes how success is evaluated. A high grade on a single task is not enough. The question becomes: Can the student identify quality? Can they seek and use feedback? Can they monitor progress? Can they justify decisions? Can they transfer standards to new contexts? Can they continue learning after formal assessment ends? These are durable capabilities.

This is where dialogic assessment links most strongly to the aims of higher education. Universities often claim to produce critical thinkers, autonomous learners and reflective professionals. These aims require assessment designs that cultivate internal judgement rather than permanent dependence on external judgement. Dialogue is not a soft alternative to standards; it is how students learn to carry standards forward.

4.3 Equity, Inclusion, and the Politics of Voice in Assessment Dialogue

Dialogic assessment carries a powerful democratic promise. By opening standards to discussion, inviting students to interpret feedback, and positioning learners as participants in judgement, it appears to challenge the secrecy and hierarchy of conventional assessment. Yet this promise cannot be accepted uncritically. Dialogue does not automatically produce equity. Indeed, if poorly designed, dialogic and participatory assessment may reproduce the very inequalities it claims to challenge. Students do not enter dialogue with equal confidence, linguistic resources, cultural familiarity, institutional trust or freedom from risk. Some know how to ask questions, challenge criteria, negotiate tasks and interpret academic expectations because they have already acquired the cultural capital of higher education. Others may experience the same invitation to dialogue as confusing, unsafe, culturally unfamiliar or inaccessible.

This section therefore confronts the equity dimension directly. Dialogic assessment can democratize assessment by surfacing hidden standards that otherwise advantage students already fluent in academic “guild knowledge.” But it can also amplify existing advantage if voice is unevenly distributed. The task is not to abandon dialogue, but to design it inclusively. Equity must be built into the Dialogic Assessment Framework from the beginning, not added later as a corrective.

1. The Promise: Making Hidden Standards Visible

The strongest equity argument for dialogic assessment begins with the hidden curriculum of assessment. Universities often present assessment as transparent because criteria, rubrics, learning outcomes and assignment briefs are published. Yet much of what students need to know remains implicit. They must learn what counts as “criticality,” “originality,” “appropriate evidence,” “coherence,” “independent thought,” “professional judgement” or “theoretical engagement.” These standards are not always fully captured in written descriptors. They are often held tacitly by disciplinary communities and recognized through experience.

Bourdieu’s (1986) analysis of cultural capital helps explain why this hiddenness matters. Higher education rewards forms of language, confidence, taste, comportment, reasoning and institutional familiarity that are unequally distributed before students arrive. Students from families or schools with experience of university often know how to interpret expectations that are never fully explained. They may know how to approach lecturers, decode feedback, perform seminar participation, use academic language, and distinguish what is formally required from what is tacitly valued. Other students may possess considerable intelligence and commitment but lack access to this academic code.

Assessment is one of the main sites where this inequality becomes consequential. A student who already understands academic conventions can read a rubric strategically, infer what the marker values and shape work accordingly. A student unfamiliar with the hidden curriculum may follow the visible instructions but miss the tacit standard. For example, they may include many sources because the criterion mentions “engagement with literature,” but fail to use those sources to construct an argument. They may describe theory because the task asks for theory, but not use theory analytically. They may interpret “participation” as speaking often, remaining silent out of respect, or waiting for the teacher’s invitation, depending on prior cultural and educational norms.

Dialogic assessment can interrupt this reproduction by making hidden standards discussable. When students examine exemplars, compare judgements, ask questions about criteria, participate in feedback tutorials and hear how teachers reason about quality, the implicit becomes more visible. Dialogue externalizes the tacit. It allows students to ask: What does “critical” look like in this discipline? Why is this paragraph considered analytical? What makes one argument stronger than another? Why does this answer meet the standard while that one does not? Such questions are equity work because they reduce the advantage of students who already know the rules.

The promise of dialogue is therefore not merely participation. It is access to the interpretive practices of the academy. Dialogic assessment opens the “guild knowledge” of academic judgement to students who might otherwise have to infer it alone. It treats standards not as private possessions of assessors but as shared objects of inquiry. This can make assessment more transparent in the deeper sense: not only visible, but understandable and usable.

2. The Peril: When Dialogue Privileges the Already Advantaged

The same dialogic practices that open standards can also reproduce inequity. Invitations to speak, question, negotiate and co-create do not affect all students equally. They may privilege students who are already confident, culturally dominant, linguistically advantaged, neurotypical, socially secure or familiar with participatory educational norms.

A feedback dialogue, for example, may benefit students who know how to ask precise questions, interpret indirect academic language and advocate for clarification. Students who feel intimidated by authority may remain silent. A co-created rubric workshop may be dominated by those who speak quickly, use confident academic vocabulary or feel entitled to challenge the teacher. A peer review session may advantage fluent speakers and disadvantage multilingual students who require more processing time. A negotiated assessment contract may reward students who already

know what kinds of tasks are valued and how to frame their interests in academically legitimate terms.

There is also a danger that “voice” becomes a progressive disguise for inequality. Institutions may celebrate student voice while hearing only certain voices. Confident students, high achievers, student representatives, native speakers, and those with time for partnership activities may become the default partners. Their experiences may then be treated as representative. Students who are working long hours, commuting, caring for family, managing disability, learning in an additional language or feeling alienated from the institution may be absent from the conversation. Participation then becomes selective, but is described as democratic.

Dialogic assessment can also create new burdens. Students may be asked to perform reflection, self-disclosure, peer critique or negotiation without adequate support. Neurodivergent students may find spontaneous discussion difficult. Students with anxiety may experience oral feedback conversations as threatening. International students may be uncertain about how much disagreement is acceptable. Students from educational cultures where teacher authority is strongly respected may experience overt critique or negotiation as disrespectful. Multilingual students may have strong ideas but less confidence in expressing them quickly in dominant academic English.

The risk is not only exclusion from dialogue, but misrecognition within it. A student who speaks hesitantly may be judged as less prepared. A student who avoids challenging peers may be seen as passive. A student who uses narrative, community-based or experiential knowledge may be judged as insufficiently academic. A student who asks for explicit instruction may be interpreted as dependent rather than as seeking access to tacit expectations. Thus, dialogue may reproduce dominant norms while appearing open.

The equity problem is therefore not solved by adding more talk. Dialogue must be structured so that participation does not depend on

already possessing the confidence, language and cultural familiarity that assessment should help develop.

3. Inclusive Design Responses

Universal Design for Learning (CAST, 2024) offers practical principles for making assessment dialogue more inclusive. Applied to dialogic assessment, UDL guides the design of multiple means of engagement, representation, and action and expression. Students should not be required to access dialogue through one mode only, such as rapid oral discussion in class. They should have multiple ways to encounter standards, contribute interpretations, ask questions, receive feedback and demonstrate evaluative judgement.

Multiple means of engagement might include small-group dialogue, anonymous question submission, peer discussion, online forums, one-to-one tutorials, written reflections and recorded explanations. Some students will participate more fully in spoken discussion; others will contribute more thoughtfully in writing or asynchronously. Multiple means of representation might include rubrics, annotated exemplars, video explanations, model feedback, visual criteria maps, bilingual glossaries where appropriate, and teacher demonstrations of judgement. Multiple means of action and expression might allow students to respond to feedback through revision notes, audio reflections, diagrams, meetings, annotated drafts or action plans.

Structured turn-taking can also reduce domination. In peer review or criteria discussion, each student may first write an individual response, then share in pairs, then contribute to the group. Roles can rotate: summarizer, questioner, evidence-finder, challenger, connector. Such structures prevent the fastest speakers from setting the entire agenda. They also give students time to process, which is particularly important for multilingual students, neurodivergent students and those less accustomed to spontaneous academic debate.

Multiple modes of participation should be normal rather than remedial. If alternative participation routes are framed as accommodations only for “special” students, they may stigmatize difference. Inclusive dialogic assessment treats variation as expected. Students can ask questions anonymously not because they are deficient, but because anonymity can surface uncertainties shared by many. Students can prepare written comments before discussion not because they cannot speak, but because thoughtful dialogue often benefits from preparation.

International and multilingual students require particular attention. Feedback and assessment dialogue often rely on idiomatic academic language. Teachers may use phrases such as “tighten the argument,” “foreground your voice,” “engage critically,” or “unpack the concept,” assuming shared understanding. Inclusive dialogue requires checking these terms, inviting students to paraphrase them, and connecting them to examples. It also requires resisting deficit views of multilingualism. Multilingual students do not lack thought because they are negotiating language; they may bring additional interpretive resources, comparative perspectives and disciplinary insights formed across educational systems.

4. Cultural Dimensions of Dialogue

Dialogue itself is culturally shaped. The dominant model of dialogic pedagogy in much higher education often assumes that good students speak openly, question authority, express individual opinions, critique peers and advocate for their needs. These behaviours align with particular Western, individualist and middle-class educational norms. They should not be universalized as the only form of active learning.

In some cultural contexts, deference to teachers is a sign of respect. Students may avoid public disagreement not because they lack critical thought, but because the social meaning of disagreement differs. In some contexts, preserving face is central to respectful interaction; direct peer critique may be experienced as humiliating. In other contexts, students

may expect the teacher to provide authoritative guidance and may interpret open-ended negotiation as lack of expertise. These differences do not mean that dialogue is impossible. They mean that dialogue must be culturally explained, negotiated and diversified.

The students-as-partners and feedback literatures warn against assuming that participation is automatically emancipatory. Partnership can impose a particular model of the ideal student: outspoken, self-directed, confident, individually expressive and comfortable with institutional ambiguity. Students who do not match this model may be seen as disengaged or resistant. A more culturally responsive approach asks how different students understand participation, authority, critique and collaboration.

This requires explicit metadialogue: dialogue about dialogue. Teachers can explain why peer critique is being used, what forms of disagreement are acceptable, how respect can be maintained, and how students may contribute in different ways. Ground rules should not simply import one cultural norm; they should be co-developed. For example, a class might agree that critique must refer to the work rather than the person, that questions may be asked in writing, that disagreement should include reasons, and that students may take time before responding.

Culturally responsive dialogic assessment also requires humility from teachers. Students' reluctance to speak may signal fear, respect, unfamiliarity, linguistic processing, previous educational experience or lack of trust. It should not be interpreted too quickly as lack of ability. The teacher's task is to create conditions where different forms of participation become possible without demanding immediate assimilation to dominant academic speech norms.

5. Voice, Asset-Based Framing and Recognition

An equity-oriented dialogic assessment framework must move beyond deficit thinking. Too often, students from marginalized backgrounds are described in terms of what they lack: lack of academic language, lack of confidence, lack of preparedness, lack of cultural capital, lack of assessment literacy. These gaps may be real in relation to dominant institutional expectations, but deficit framing obscures the resources students bring.

Yosso's (2005) concept of community cultural wealth offers an important corrective. Developed as a critical response to deficit interpretations of cultural capital, community cultural wealth identifies six overlapping forms of capital: aspirational, linguistic, familial, social, navigational and resistant. Applied to assessment, this perspective asks how students' existing resources can be recognized and mobilized rather than ignored.

Dialogic assessment can support asset-based practice when it treats students as knowers. A student's community experience may enrich analysis. Multilingual competence may provide comparative insight. Navigating unfamiliar systems may demonstrate resilience and strategic intelligence. Family and community responsibilities may shape forms of time management and ethical awareness. Prior educational traditions may offer alternative understandings of respect, knowledge and collaboration. These resources should not replace disciplinary standards, but they can enter dialogue with them.

Recognition is crucial here. As argued in the previous section, assessment is identity-shaping. Students must be recognized not only as learners with deficits to be corrected but as developing knowers with resources to contribute. This does not mean romanticizing all prior knowledge or avoiding critique. It means that critique begins from the assumption that the student is capable of meaningful participation in knowledge-making.

Voice, then, is not simply speaking. Voice is being heard as someone whose contribution can matter. A student may speak, but if their contribution is translated immediately into a deficiency, voice has not been recognized. Conversely, a student may contribute through writing, peer support, reflective synthesis or careful questioning. Dialogic assessment should expand what counts as voice.

Asset-based dialogue also changes feedback. Instead of saying only, “Your writing lacks academic style,” a teacher might say, “Your example from community practice is powerful; now we need to connect it more explicitly to the theoretical claim so that its analytical value becomes visible.” The student’s resource is recognized, and the disciplinary demand is maintained. This is equity without lowered standards.

4.4 An Integrated Model: The Dialogic Assessment Framework (DAF)

The preceding sections have examined the mechanisms that make dialogic assessment possible and the conditions that make it fragile. Section 4.1 argued that assessment dialogue works through relational and affective mechanisms: trust, safety, recognition, vulnerability and accountable power. Section 4.2 showed how dialogue produces durable learning through the internalization of standards, self-regulation, comparison, calibration and sustainable evaluative judgement. Section 4.3 confronted the equity dimension, arguing that dialogue can democratize assessment only when participation is structured, inclusive and asset based. Chapter 3 provided illustrative designs for these arguments: dialogic feedback in disciplinary writing, co-created criteria and negotiated assessment, programme-level portfolios and programmatic assessment, and peer-based dialogue.

This section synthesizes those arguments into the Dialogic Assessment Framework (DAF). The DAF is this book’s proposed conceptual synthesis. It has not been empirically validated as a framework. It offers a

multidimensional model for understanding, designing and researching assessment as dialogue. It does not propose a single method, tool or technique. Rather, it identifies the conditions, moves and trajectories through which assessment can shift from one-way measurement toward shared and ultimately internalized judgement.

The framework begins with a simple premise: assessment becomes dialogic when students are enabled to participate meaningfully in the interpretation, use and development of standards over time. But this participation is not merely conversational. It is relational, structural, epistemic, developmental and political-ethical. A feedback tutorial, peer review exercise or co-created rubric is not dialogic simply because students speak. It becomes dialogic only when the exchange is connected to trust, standards, uptake, power and learning across time.

1. The Five Dimensions of Dialogic Assessment

The first dimension is relational. Dialogue depends on trust, recognition and psychological safety. Students need to believe that assessment interaction is fair, meaningful and oriented toward development. They must be able to disclose misunderstandings, ask for clarification, receive critique and revise self-understanding without excessive threat. The relational dimension does not eliminate challenge; it makes challenge usable. It is visible in draft tutorials where students can explain intentions, in mentoring relationships where feedback accumulates across time, and in peer review groups where students learn to critique work without attacking persons.

Recognition is central within this dimension. Assessment is not only a judgement of work; it is also a response to a developing learner. Dialogic assessment recognizes students as emerging participants in a disciplinary or professional community. It addresses them as capable of judgement, not merely as objects of judgement. Without recognition, dialogue becomes procedural. With recognition, critique can become identity-

forming in a positive sense: students begin to see themselves as writers, researchers, practitioners or professionals in development.

The second dimension is structural. Dialogue requires assessment architectures that create loops: work, feedback, interpretation, revision and reuse. A single comment after a terminal assignment may be informative, but it does not constitute a dialogic system. Structural dialogue appears in two-stage assignments, feed-forward designs, portfolio sequences, patchwork texts and programmatic assessment. These designs make time available for feedback to matter. They connect tasks so that students can act on assessment information rather than merely receive it.

The structural dimension also includes workload, timetabling, staffing and digital infrastructure. Dialogue cannot be sustained by teacher goodwill alone. If assessment calendars cluster deadlines, if feedback arrives after modules have ended, if tutors have no time for conversation, or if data is fragmented across platforms, dialogic assessment collapses into aspiration. Structural design determines whether dialogue is feasible.

The third dimension is epistemic. This concerns how students come to understand what counts as quality. Standards are not self-evident. They are learned through exemplar work, comparison, criteria discussion, peer judgement, moderation exercises and feedback interpretation. The epistemic dimension therefore concerns the co-construction and clarification of standards. It asks how students are given access to the tacit knowledge of a discipline.

The illustrative designs in Chapter 3 show this dimension clearly. In disciplinary writing, students learn what counts as argument or analysis through discussion of drafts and exemplars. In co-created rubric exercises, students articulate and refine criteria. In peer review, students test judgements against others' interpretations. In programmatic assessment, students and mentors interpret evidence against evolving competence standards. Across these cases, knowledge of quality is not transmitted as a finished object. It is constructed through participation in judgement.

The fourth dimension is developmental. Dialogic assessment aims beyond immediate performance improvement. Its long-term purpose is the development of evaluative judgement, self-regulation and internal feedback. Students should gradually become able to judge their own work and the work of others without continuous teacher intervention. Dialogue is therefore transitional: it supports the movement from external judgement to internalized judgement.

This dimension connects directly to Vygotskian internalization. Students first encounter judgement socially, through teachers, peers, mentors, criteria and exemplars. Through repeated participation, they appropriate evaluative language and standards. Eventually, the questions once asked by teachers become questions students ask themselves. The developmental dimension is visible when students begin to request specific feedback, justify revisions, compare their work with exemplars, evaluate peer work, and set goals based on prior performance.

The fifth dimension is political-ethical. Assessment is always shaped by power. Teachers and institutions define tasks, set standards, award marks and certify competence. Dialogue does not abolish this asymmetry. Instead, dialogic assessment seeks to make power visible, accountable and educative. Students should know what is negotiable, what is fixed, why standards matter and how judgements are made.

The political-ethical dimension also includes equity and voice. Dialogue can democratize assessment by surfacing hidden standards, but it can also privilege students already confident in academic culture. Therefore, dialogic assessment must be inclusive by design. It must structure participation, provide multiple modes of voice, attend to linguistic and cultural diversity, and recognize students' community cultural wealth. Without this dimension, dialogic practices risk becoming dialogue-washing: progressive in language but unequal in effect.

2. Dialogic Moves: A Vocabulary for Design and Research

The DAF identifies a set of dialogic moves that operationalize these dimensions. These moves are not techniques in isolation; they are actions through which assessment dialogue is enacted.

The first move is eliciting. Eliciting invites students' interpretations, questions, intentions and self-evaluations into the assessment process. Feedback request sheets, draft cover notes, self-assessment prompts and mentoring conversations all perform this move. Eliciting changes the student's role from recipient to contributor.

The second move is exemplifying. Exemplifying makes standards visible through concrete cases. Annotated exemplars, sample essays, model solutions, portfolio extracts and contrasting performances help students see quality in action. Exemplifying is epistemic because it translates abstract criteria into situated judgement.

The third move is comparing. Comparing asks students to place work alongside reference information: exemplars, criteria, peer work, previous attempts or professional standards. This is the core cognitive move of internal feedback. It enables students to notice gaps, possibilities and patterns.

The fourth move is articulating. Students must put judgement into words. They explain why one exemplar is stronger, justify peer feedback, describe how they used comments, or write a reflective commentary. Articulation makes evaluative reasoning visible and available for refinement.

The fifth move is negotiating. Negotiating involves discussion of criteria, task formats, standards, evidence or feedback meanings. It does not mean that all standards are optional. It means that assessment meanings are opened to reason-giving and bounded participation.

The sixth move is calibrating. Students compare their judgements with those of peers, teachers, mentors or benchmark standards. Calibration helps students refine self-evaluation and reduce novice misjudgement.

Peer review, comparative judgement and moderation workshops all rely on this move.

The seventh move is recognizing. Recognizing addresses the student as a developing knower. It acknowledges effort, identity, prior knowledge and emerging capability while maintaining standards. Recognition is especially important for equity because it resists deficit framing.

The eighth move is feeding-forward. Feeding-forward connects feedback to future action. It requires revision, transfer, planning or reuse. A comment that cannot be acted upon remains weakly dialogic. Feeding-forward completes the feedback loop.

The ninth move is documenting. Dialogic assessment often requires records: action plans, portfolio reflections, peer-review logs, mentoring notes or revision commentaries. Documentation supports continuity and accountability, but must not replace dialogue.

Together, these moves provide a shared vocabulary for design and research. They allow us to ask not simply whether an assessment is “dialogic,” but which dialogic moves it contains, which dimensions they activate, and what conditions support or limit them.

3. Trajectories of Assessment Dialogue

The DAF also identifies four trajectory types through which assessment systems may develop.

The first trajectory is monologic measurement. Here assessment is organized as one-way judgement. Students submit work; teachers grade it; feedback, if provided, explains the mark. The student’s role is mainly to comply and receive. The baseline comment-and-grade essay cycle described in Chapter 3 illustrates this trajectory. It may produce grades, but it provides weak conditions for uptake, internalization or shared standards.

The second trajectory is responsive feedback. In this form, feedback remains largely teacher-generated, but students are invited to respond.

They may ask questions, complete feedback reflection sheets, attend optional tutorials or write action plans. This is a significant improvement over monologic assessment because it acknowledges uptake. However, standards and judgement remain mainly teacher-held. Many two-stage writing assignments begin here: students receive feedback and revise, but the teacher still controls the evaluative frame.

The third trajectory is co-constructed judgement. Here students participate in interpreting and sometimes shaping standards. Co-created criteria, exemplar workshops, marking exercises, peer review and negotiated assessment contracts belong to this trajectory. Students are not merely acting on feedback; they are learning how judgement is made. Chapter 3's cases of co-created rubrics and peer-based dialogue exemplify this trajectory.

The fourth trajectory is sustainable or internalized judgement. This is a developmental aim, not a final state of self-sufficiency. Students become increasingly capable of judging their own work, seeking useful feedback, comparing performance against standards, regulating learning and justifying decisions. Programme-level portfolios, ipsative feedback and programmatic assessment aim toward this trajectory. The student becomes less dependent on single teacher comments because evaluative judgement has been internalized.

These trajectories are not rigid stages through which every institution must pass. Assessment systems may contain multiple trajectories simultaneously. A programme may use monologic final exams, responsive essay feedback, co-created criteria in one module and portfolio mentoring elsewhere. The value of the model is diagnostic. It reveals where dialogue is located, where it is absent, and whether the developmental endpoint is being designed for or merely assumed.

4. Surface Features and Deep Conditions

A central distinction in the DAF is between surface dialogic features and deep dialogic conditions. Surface features are visible techniques: peer review, feedback conversations, co-created rubrics, reflection sheets, student voice panels, online discussion boards, portfolio meetings. These are important, but they do not guarantee dialogic assessment.

Deep conditions are the relational, structural and political-ethical arrangements that make techniques meaningful. A peer review activity without training, trust or revision opportunity may produce superficial comments. A feedback tutorial in a high-stakes setting without psychological safety may produce defensiveness. A co-created rubric without genuine influence may become tokenism. A portfolio without mentoring may become bureaucracy. A student voice process that hears only confident students may reproduce inequality.

This is the problem of dialogue-washing. Dialogue-washing occurs when institutions adopt the language or surface techniques of dialogue while leaving the underlying assessment relation unchanged. Students appear to participate, but power remains hidden. Feedback appears dialogic, but no opportunity for uptake exists. Criteria appear co-created, but the teacher's decisions are predetermined. Partnership appears inclusive, but only already-privileged students are involved.

The DAF guards against dialogue-washing by asking five questions. First, is there relational safety for honest participation? Second, is there a structural loop that allows feedback to be used? Third, are standards made available through concrete epistemic work? Fourth, does the design build long-term evaluative judgement? Fifth, are power and equity explicitly addressed? If the answer to these questions is no, then the practice may have dialogic features but not dialogic depth.

5. Mapping Chapter 3 Illustrative Designs

The disciplinary writing case illustrates the movement from responsive feedback to co-constructed and internalized judgement. Two-stage assignments, draft tutorials, feedback request sheets and exemplar discussions activate eliciting, exemplifying, comparing, articulating and feeding-forward. The relational dimension is present when students trust the tutor enough to discuss unfinished writing. The epistemic dimension is present when students learn what counts as argument, evidence and criticality. The developmental dimension appears when students begin to revise and later self-monitor their writing.

The co-created assessment case illustrates the epistemic and political-ethical dimensions most strongly. Students participate in criteria design, rubric discussion, negotiated tasks and moderation exercises. The key moves are negotiating, articulating, calibrating and recognizing. The promise is that students gain ownership and understand standards more deeply. The risk is unequal participation or cosmetic sharing of authority. This case shows why DAF must include power and equity, not only learning design.

The programme-level case illustrates the structural and developmental dimensions. Portfolios, patchwork texts, programmatic assessment and ipsative feedback extend dialogue across time. The key moves are documenting, feeding-forward, comparing and articulating, supported by mentoring. These designs support continuity and self-regulation by allowing students to interpret evidence of development. Their risks include workload, bureaucracy, reliability concerns and unequal mentoring quality. They show that dialogic assessment requires infrastructure.

The peer-based dialogue case illustrates scalable epistemic and developmental practice. Peer review, adaptive comparative judgement, calibrated peer review and group assessment distribute judgement among students. The key moves are comparing, articulating, calibrating and recognizing. Peer dialogue can develop evaluative judgement because students learn by judging others' work. But it requires structured talk, training, trust and attention to intercultural dynamics. It shows that dialogue at scale is possible, but not automatic.

Together, these illustrative designs suggest that no single practice embodies the full DAF. Each combines dimensions and moves differently. The framework is therefore not a checklist of required techniques. It is an analytical lens for understanding how assessment designs create, limit or deepen dialogue.

Beyond the Rubric in Practice

Implications for Curriculum, Faculty, Policy, Technology, and Equity

5.1 Redesigning Curriculum and Assessment Regimes for Dialogue

If dialogic assessment remains confined to individual tasks, it will remain fragile. A single teacher may redesign an essay assignment around drafts and feedback dialogue; another may co-create criteria with students; a third may introduce peer review or reflective portfolios. These innovations can improve learning locally. Yet when the wider programme continues to operate through fragmented modules, dense summative assessment, terminal feedback and isolated grading events, the dialogic innovation is absorbed back into a monologic regime. Students may experience one module as dialogic and developmental, while the rest of the programme teaches them that assessment is primarily about producing graded products, moving on quickly, and treating feedback as an afterthought.

The implication is clear: dialogic assessment requires programme-level architecture. The unit of redesign cannot be only the assignment, or even the module. It must be the programme as a developmental pathway. Students need repeated, sequenced and coherent opportunities to encounter standards, practise judgement, receive and use feedback,

compare work, revise strategies and internalize evaluative capacity. A programme that claims to value feedback but does not create time for uptake is not designed for dialogue. A programme that claims to cultivate independent judgement but keeps all judgement in the hands of markers until final year is not designed for development. A programme that claims to promote equity but leaves students to decode different assessment expectations in every module is not designed for inclusion.

1. From Task-Level Fixes to Programme-Level Architecture

The most common form of assessment reform is task-level innovation. A lecturer changes an essay into a portfolio, introduces peer feedback, adds a formative draft, uses audio comments or creates a rubric workshop. Such practices matter, but they are insufficient when surrounded by assessment systems that remain unchanged. Students learn from the whole assessment ecology, not only from the most innovative task within it. If most modules reward last-minute performance, if feedback arrives too late to be reused, if grades dominate attention, and if standards vary without explanation, then the underlying message remains monologic: perform, receive judgement, move on.

Programme-level assessment design begins by mapping the curriculum as an assessment journey. Rather than asking only, “What does this module assess?” programme teams ask: What capabilities should students develop across the degree? Where are these introduced, practised, assessed and revisited? How do students receive feedback on them? Where do they use that feedback? Where do they practise judging quality? Where do assessment demands cluster? Where are students over-assessed? Where are they under-supported?

The TESTA project (Transforming the Experience of Students through Assessment), including the work of Jessop and Tomas (2017), is important because it shifts attention from individual modules to programme-level assessment patterns. Its contribution lies not merely in improving feedback techniques, but in making visible the assessment

regime as a whole: the number of tasks, timing of deadlines, distribution of formative and summative work, variety of assessment types, feedback opportunities, and coherence across modules. Such mapping often reveals problems that no individual teacher can solve alone. Students may face multiple deadlines in the same week, receive repeated feedback on similar issues without structured reuse, encounter inconsistent language criteria, or complete too many small summative tasks that leave little time for dialogue.

Programme-wide mapping therefore functions as the diagnostic phase of dialogic redesign. It identifies where the programme creates loops and where it creates dead ends. A loop exists when students produce work, receive or generate feedback, interpret it, and apply it to later work. A dead end exists when feedback is returned after the task has no future, or when the next task bears no meaningful relation to the previous one. The purpose of mapping is to replace accidental assessment accumulation with intentional developmental sequencing.

2. Building Feedback Loops into the Curriculum

Dialogic assessment depends on loops. At programme level, this means designing linked and nested tasks that allow feedback to be reused. A first-year short analytical response might feed into a longer essay. The essay might feed into a research proposal. The proposal might feed into a methods plan. The methods plan might feed into a final-year dissertation or capstone project. The point is not simply that tasks follow one another chronologically, but that each task creates feedback and judgement that students must use later.

Nested assessment reduces the waste of feedback. In conventional modular systems, a student may receive feedback on essay structure in one module and then write an entirely different kind of assessment in another. The feedback may still be relevant, but the programme does not help the student transfer it. In a nested design, transfer is built into the task sequence. Students may be required to identify which previous feedback

they used, show how they revised a strategy, or reflect on recurring patterns across assignments.

This requires what may be called “slow assessment.” Slow assessment does not mean less rigour or lower expectations. It means fewer isolated assessment events and more time for students to produce, receive, discuss, revise and integrate. It resists the culture of constant submission in which students move from deadline to deadline without developmental consolidation. Slow assessment values the maturation of judgement over the multiplication of graded outputs.

Reducing summative density is essential. Many programmes are assessment-heavy but feedback-poor. Students complete numerous graded tasks, but few of those tasks create meaningful opportunities for dialogue or reuse. Summative overload consumes staff marking time and student energy. It also increases strategic behaviour: students focus on marks, not learning. A dialogic programme may therefore need fewer summative tasks, but better-designed ones. Some tasks should be decoupled from grades or given low stakes so that students can use them for practice, experimentation and feedback.

Decoupling does not mean removing accountability. Low-stakes tasks can be required, structured and meaningful without carrying heavy grade weight. A formative literature-map exercise, peer-reviewed draft, self-assessment commentary or exemplar-analysis task can prepare students for summative work while reducing the fear attached to every performance. This creates room for candour. Students are more likely to disclose misunderstanding and act on critique when not every weakness is immediately converted into a mark.

3. Programmatic and Integrative Designs at Scale

The Chapter 3 discussion of portfolios, patchwork texts and programmatic assessment offers programme-level resources for dialogic redesign. These designs treat assessment as a longitudinal system rather than a series of disconnected events. Their common feature is the accumulation and interpretation of evidence over time.

Portfolios can serve as integrative spaces where students collect evidence of learning, reflect on feedback, identify development and construct a narrative of progress. Used well, the portfolio becomes a dialogue between the student, mentor and evolving body of evidence. It helps students see patterns obscured by individual module marks. For example, a sociology student might track development in argumentation, theory use, research design, ethical reasoning and public communication across the programme. The portfolio then becomes both an assessment tool and a self-regulation tool.

Patchwork assessment can operate similarly. Students create smaller pieces of work across a module or programme and later integrate them through a reflective synthesis. This structure allows feedback to be distributed over time and encourages students to make connections. It is particularly useful where learning is cumulative, interdisciplinary or practice-based. It also allows students to see their thinking developing, rather than treating the final submission as the only evidence that matters.

Programmatic assessment offers a more systematic model, especially from medical education. Its key principle is that many low-stakes data points generate feedback, while high-stakes decisions are made holistically across accumulated evidence. This model is powerful because it changes the role of individual assessments. Each assessment moment becomes an opportunity to learn, not a terminal verdict. Students engage with evidence over time, often with a coach or mentor, and final decisions are based on patterns rather than isolated performances.

Adapting programmatic assessment beyond medical education requires care. Not every discipline needs the same infrastructure as health professions education. Yet the principles transfer: multiple evidence points, feedback-rich low-stakes assessment, longitudinal advising, evidence aggregation, and holistic review at key progression points. A humanities programme might not need workplace-based competence data, but it can still collect evidence of writing development, interpretive judgement, research skill and reflective capacity across time. An engineering programme can collect design iterations, lab reports, peer evaluations, project milestones and professional reflections. A business programme can gather case analyses, presentations, team contributions, simulations and reflective decisions.

Such designs require data and advising infrastructure. Feedback must be retrievable. Students and advisors need access to prior work, comments, action plans, self-assessments and development goals. If assessment evidence is scattered across separate modules and platforms, longitudinal dialogue becomes difficult. A programme-level portfolio system, advising record or feedback dashboard can help, but only if designed for learning rather than surveillance. Data must support interpretation, not merely accumulate.

Advising is equally important. Students often cannot make sense of longitudinal evidence alone. They need mentors, tutors or academic advisors who help them identify patterns, set goals and connect feedback across modules. This role should be recognized as part of assessment architecture, not treated as optional pastoral work. Without advising, integrative assessment can become paperwork. With advising, it can become developmental dialogue.

4. Workload and Sustainability

Programme-level dialogic assessment must be realistic about time. It cannot simply add dialogue on top of existing assessment loads. If staff are already overburdened by marking, moderation, administration and student support, asking them to provide more tutorials, more comments and more reflective feedback will fail. Sustainable redesign requires deciding what to stop doing.

The first activity to reduce is repetitive teacher commenting on terminal tasks. If students cannot use feedback, extensive comments may have low educational return. Programmes should shift effort from late explanatory feedback to earlier feed-forward, exemplar dialogue, peer review and targeted feedback on work in progress. This may reduce the quantity of teacher comments while increasing their usefulness.

The second activity to reduce is unnecessary summative assessment. Not every learning activity needs a grade. Excessive grading produces workload for staff and anxiety for students. A programme should identify which tasks genuinely need summative judgement and which can function as formative practice, peer dialogue or self-assessment.

The third activity to reduce is duplicated assessment of the same capability without developmental connection. If several modules assess essay writing but none connect feedback across tasks, staff may repeatedly comment on the same weaknesses. Programme teams can coordinate so that capabilities are introduced, practised and assessed progressively.

Strategic placement of high-touch dialogic moments is essential. Intensive teacher-student dialogue should be reserved for points where it has the greatest developmental value: transition into university, first major disciplinary writing task, research proposal, placement preparation, dissertation planning, remediation after failure, or capstone integration. Other points can use lower-touch but still dialogic designs: peer review, group feedback, annotated exemplars, self-assessment prompts, digital discussion and feedback response tasks.

Student workload also matters. Dialogic assessment asks students to do more than submit work. They must read exemplars, review peers, reflect on feedback, revise drafts and maintain portfolios. These activities require time and must be built into credit-bearing workload. If they are added informally, students may treat them as optional or burdensome. Programme design should make the work of dialogue visible and legitimate.

5. Change Management and Reform Sequencing

Programme-level redesign is not only a pedagogical challenge; it is a change-management challenge. Assessment practices are tied to disciplinary identities, workload habits, quality assurance requirements, student expectations and institutional regulations. Reform can generate fatigue if staff experience it as another imposed initiative. It can also fail if students do not understand the new assessment culture.

A sensible reform sequence begins with mapping rather than immediate redesign. Programme teams should first map assessment tasks, deadlines, feedback types, criteria language, formative opportunities, summative density and capability progression. This creates shared evidence and reduces blame. The question becomes not “Which lecturer is doing assessment badly?” but “What pattern are we collectively creating for students?”

The second stage is identifying pressure points. These might include deadline bunching, over-assessment, lack of formative practice, inconsistent feedback language, weak final-year preparation, or poor student engagement with comments. Reform should begin with problems staff and students already recognize. This builds legitimacy.

The third stage is selecting a small number of programme-level design principles. For example: every major summative task should be preceded by a formative comparison activity; feedback on one task should be explicitly reused in another; first-year students should receive structured assessment literacy teaching; every year should include at least one self-

assessment or peer-review activity; final-year capstone work should include longitudinal advising.

The fourth stage is piloting and iterating. Whole-programme reform need not mean changing everything at once. A programme might redesign first-year assessment first, then introduce portfolios in second year, then revise capstone advising. Iteration reduces risk and allows staff to learn from implementation.

The fifth stage is faculty development and student induction. Staff need shared language around feedback, criteria, dialogue, calibration and evaluative judgement. Students need to understand why assessment is changing. If students expect every task to be graded and every feedback comment to come from the teacher, they may resist peer review or low-stakes formative work. The programme must explain the rationale: assessment is being redesigned to build judgement, not to reduce teacher responsibility.

The sixth stage is evaluation. Programmes should monitor not only grades but feedback uptake, student workload, staff workload, equity of access, assessment literacy and progression of evaluative judgement. Reform should be evidence-informed but not paralysed by the absence of perfect metrics.

5.2 Faculty Development and Assessment Literacy: Preparing Educators for Dialogue

The Dialogic Assessment Framework cannot be enacted by design documents alone. Even the most coherent programme architecture will fail if educators continue to understand assessment as the delivery of judgement and feedback as the transmission of comments. Dialogic assessment lives or dies in teacher capability, belief, institutional support and professional identity. It requires educators who can design feedback loops, facilitate assessment dialogue, interpret standards with students, manage affect, work with uncertainty, calibrate judgement with colleagues, and share authority without abandoning academic responsibility.

This section therefore focuses on the educators who must enact dialogic assessment. The central argument is that implementing the DAF requires a new form of educator assessment literacy: one that goes beyond rubric-writing, marking consistency and feedback technique. Educators need relational competence, design competence, epistemic competence, equity competence and political judgement. They must know not only how to produce feedback, but how to create conditions under which students can use feedback, participate in judgement and develop evaluative capacity. This is a demanding professional shift. It involves conceptual change, not just skill acquisition.

1. Teacher Feedback and Assessment Literacy

Teacher feedback literacy has emerged as a key construct in higher education assessment scholarship. Carless and Winstone (2023) define it as the knowledge, expertise and dispositions educators need to design feedback processes that enable student uptake. This definition is important because it moves feedback literacy beyond the ability to write better comments. The literate feedback teacher is not simply a clear communicator. They are a designer of feedback ecologies.

Building on this literature, the DAF proposes five interrelated capacities. The first is design capacity. Educators must know how to design tasks so that feedback can be used. This involves sequencing assignments, building opportunities for revision, connecting feedback across tasks, embedding peer dialogue, and ensuring that comments have a future destination. A teacher who writes excellent feedback after a terminal task has not completed the feedback process if students have no meaningful opportunity to act on it.

The second capacity is epistemic. Educators must understand how students come to grasp standards. Criteria and rubrics matter, but they do not speak for themselves. Teachers need to use exemplars, comparative judgement, moderation exercises and dialogue to make tacit disciplinary standards visible. They must be able to explain why a piece of work is strong, not merely state that it is. They must help students see quality in context.

The third capacity is relational. Dialogic assessment requires trust, recognition and psychological safety. Teachers must know how to frame critique developmentally, invite questions without humiliation, respond to emotional reactions, and maintain standards while recognizing the student as a developing knower. Feedback literacy therefore includes affective and interpersonal judgement. A technically accurate comment may fail if delivered in a relation of fear, mistrust or indifference.

The fourth capacity is pragmatic. Educators work under time constraints, large cohorts, administrative systems and institutional deadlines. Teacher feedback literacy therefore includes knowing how to make feedback sustainable: when to use group feedback, when to use peer review, when to provide individual response, when to use audio or screencast feedback, and when to stop writing comments that students cannot use. Pragmatic literacy is not a lowering of ambition; it is the condition for sustaining dialogic practice.

The fifth capacity is equity-oriented. Educators must design assessment dialogue so that it does not privilege only confident, culturally

dominant or linguistically advantaged students. This means structuring participation, offering multiple modes of response, attending to multilingual students, recognizing diverse forms of knowledge and monitoring whose voices are heard. Inclusive dialogic assessment is not an optional extension of feedback literacy; it is part of its core.

Boud and Dawson's (2023) multi-level framing of teacher feedback literacy is useful because it locates competence at macro, meso and micro levels. At the micro level, teachers respond to student work. At the meso level, they design modules and tasks. At the macro level, they contribute to programme assessment systems. Dialogic assessment requires literacy at all three levels. A teacher who is skilled in face-to-face feedback but isolated in a poorly sequenced programme cannot fully enact the DAF. Conversely, a programme-level framework will not work if individual teachers lack the relational and epistemic competence to enact it.

2. Conceptual Change: From Transmission to Dialogue

Professional development in assessment often fails because it treats practice as a technical problem. Teachers are offered workshops on rubric design, feedback efficiency or moderation procedures, but the deeper conception of assessment remains unchanged. If educators continue to believe that feedback is information teachers give to students, then dialogic techniques will be absorbed into a transmission model. Peer review becomes a way to generate more comments. Rubrics become more detailed instructions. Feedback request sheets become administrative forms. Dialogue becomes surface activity.

The shift required is conceptual. Educators must move from a transmission or measurement conception to a dialogic and process conception. In the transmission model, the teacher's main responsibility is to judge performance accurately and communicate deficiencies clearly. The student's responsibility is to receive and apply the information. In the dialogic model, the teacher's responsibility is to design and facilitate

processes through which students make sense of standards, generate internal feedback, act on critique and develop evaluative judgement.

Beliefs drive practice. An educator who believes students are passive recipients may interpret poor feedback uptake as student laziness. An educator who believes feedback is a process may ask whether the design allowed students to interpret and reuse feedback. An educator who sees assessment primarily as certification may prioritize grading efficiency. An educator who sees assessment as developmental may ask how certification can be embedded within learning. These beliefs shape the choice of tasks, the timing of feedback, the tone of comments, the willingness to involve students, and the interpretation of student resistance.

Conceptual change is difficult because assessment beliefs are embedded in academic identity. Many educators were themselves trained in systems where expertise meant authoritative judgement. Their legitimacy may feel tied to being the person who knows, evaluates and decides. Dialogic assessment can feel like a threat: a loss of control, dilution of standards, or excessive accommodation of students. Professional development must therefore address fears and values, not only skills. It must show that dialogue does not mean abandoning standards. It means making standards learnable.

Conceptual change also requires evidence from practice. Teachers are more likely to shift beliefs when they see that students can produce better questions, use feedback more productively, and improve judgement through structured dialogue. Small design experiments can be powerful: a teacher introduces feedback request sheets and discovers that students' questions reveal misunderstandings previously hidden; a moderation workshop shows staff how different criteria are interpreted; a peer review exercise demonstrates that giving feedback helps students revise their own work. Such experiences can unsettle transmission assumptions.

3. Development Approaches: Modelling Dialogue in Professional Learning

If educators are to enact dialogic assessment, their professional development should itself be dialogic. It should not merely transmit principles of dialogue through lectures or policy documents. Academic development must model the practices it advocates: eliciting experience, examining evidence, comparing examples, discussing standards, negotiating meanings and reflecting on action.

One powerful approach is the exemplar and moderation workshop. Staff bring samples of student work, apply criteria, compare judgements and discuss differences. Such workshops serve several purposes. They improve consistency, but more importantly they surface tacit standards. Staff discover that they may use the same criteria language while valuing different features of work. They must articulate what they mean by criticality, originality, coherence or professional judgement. This mirrors the student learning process and helps educators understand why students struggle with criteria.

Calibration workshops can also support teacher feedback literacy. Educators review sample feedback comments and discuss which are most likely to support uptake. They examine whether comments are actionable, whether they explain standards, whether they invite student response, and whether they connect to future work. This shifts attention from feedback quantity to feedback function.

Communities of practice provide a longer-term structure. Dialogic assessment cannot be learned fully in a one-off workshop. Educators need ongoing spaces to share designs, troubleshoot problems, examine student responses and refine practice. A community of practice might focus on peer review, programme-level feedback loops, inclusive assessment dialogue, or portfolio advising. Its value lies in sustained collective inquiry rather than compliance training.

Student partners can play a significant role in staff development. Students can help educators understand how feedback is experienced, where criteria remain obscure, which comments are usable, and how assessment practices affect motivation and belonging. Involving students in staff workshops can disrupt assumptions. Teachers may believe their feedback is clear; students may reveal that the language is difficult to interpret. Teachers may believe optional feedback opportunities are equitable; students may explain that only confident students attend. Student partnership makes professional learning more dialogic and more accountable.

However, student involvement must be carefully designed. It should not rely only on high-achieving or highly confident students. It should include diverse voices and avoid placing students in risky positions where they must criticize staff without protection. Student partners should be prepared, supported and recognized for their labour.

4. The Casualization Challenge

The move toward dialogic assessment is complicated by casualization. Many higher education systems rely heavily on sessional, adjunct or hourly paid staff to teach, mark and provide feedback. These colleagues often bring expertise, commitment and valuable professional experience. The problem is not casual teachers themselves. The problem is the institutional structure that gives them limited time, limited security, limited access to professional development and limited continuity with students.

Dialogic assessment depends on continuity of relationship and shared understanding of standards. Casualized marking can undermine both. A sessional marker may not have designed the task, participated in the teaching, or joined moderation discussions. They may be paid only for marking time, not for feedback dialogue, meetings or student consultation. They may not see students again after returning comments.

Under such conditions, feedback is pushed toward efficient judgement rather than developmental conversation.

Casualization also affects consistency. If large numbers of markers interpret criteria differently, students receive uneven messages about quality. Moderation can reduce this, but only if sessional staff are included meaningfully in calibration, not merely sent a rubric. A marker who is excluded from the programme's assessment culture cannot be expected to enact its dialogic principles.

There are institutional responses. First, sessional staff must be included in assessment design briefings, moderation workshops and feedback development activities, with payment for that time. Second, feedback roles should be clarified: what kind of comments are expected, how students will use them, what future task they connect to, and what dialogue opportunities exist. Third, programmes should create shared banks of annotated exemplars and feedback principles so that all markers work from a common interpretive base. Fourth, where possible, sessional teachers should have continuity with cohorts, not be used only as anonymous markers. Fifth, online professional development modules can support flexible access, but they should not replace live calibration and community.

Institutions must recognize that casualized labour is not merely an employment issue; it is an assessment quality issue. If the university wants dialogic feedback, it must resource the people expected to provide it. Otherwise, the DAF becomes another demand placed on those with the least institutional security.

5. A Capability Framework for Dialogic Assessment Literacy

The following capability framework summarizes what educators need to enact dialogic assessment.

First, educators need assessment design capability. They can sequence tasks, create feedback loops, reduce terminal feedback, and design opportunities for students to use feedback.

Second, they need standards-explication capability. They can use exemplars, criteria discussion, comparative judgement and moderation to make tacit standards visible.

Third, they need dialogic facilitation capability. They can lead feedback conversations, peer review sessions, rubric discussions and reflective activities in ways that elicit student thinking and support participation.

Fourth, they need relational capability. They can build trust, frame critique developmentally, manage affect and recognize students as developing knowers.

Fifth, they need evaluative judgement capability. They can make and justify judgements, calibrate with colleagues, and help students learn to judge.

Sixth, they need equity capability. They can identify hidden curriculum effects, structure voice inclusively, support multilingual and diverse learners, and use asset-based feedback.

Seventh, they need pragmatic sustainability capability. They can decide where high-touch feedback is necessary, where peer or group dialogue will suffice, and what low-value practices should be stopped.

Eighth, they need programme-level capability. They can contribute to assessment mapping, shared criteria development, feedback sequencing and longitudinal student development.

This framework deliberately combines technical, relational, epistemic and political dimensions. Dialogic assessment literacy is not one skill. It is a professional orientation supported by a repertoire of practices.

5.3 Quality, Standards, and Policy: Reconciling Dialogue with Accountability

Dialogic assessment faces its hardest institutional test at the level of policy and quality assurance. The framework developed in this book argues that standards are learned through dialogue, that judgement is socially mediated, and that students should participate in making assessment meanings explicit. Yet universities operate within accountability systems that demand comparability, reliability, auditability and public confidence. External examiners, regulatory agencies, professional bodies, league tables, appeals procedures and institutional quality codes all exert pressure toward standardization. The question is therefore unavoidable: can assessment be dialogic and still be defensible?

This section argues that the tension is real, but often misdescribed. The problem is not that dialogue threatens standards while quality assurance protects them. Rather, standards in higher education are already social, interpretive and dialogic. Reliability is not achieved simply by writing more detailed criteria or enforcing rubric compliance. It can be supported through professional communities developing shared judgement through calibration, moderation, comparison and reasoned deliberation, alongside checks on the consistency and validity of decisions. The policy task, then, is not to suppress dialogue in the name of assurance, but to institutionalize the right forms of dialogue as part of assurance.

1. The Accountability–Dialogue Tension

Quality assurance regimes are designed to protect public trust in higher education. They ask institutions to demonstrate that awards are credible, that students are treated fairly, that standards are comparable across cohorts, and that assessment decisions can be defended. These are legitimate concerns. A university degree has social value because others trust that it represents a certain level of achievement. Assessment must therefore be accountable beyond the immediate teacher-student relationship.

However, accountability systems often push assessment toward defensiveness. Institutions respond to external scrutiny by producing more documentation: rubrics, grade descriptors, moderation forms, external examiner reports, audit trails, assessment policies and standard operating procedures. These documents are intended to make judgement transparent and consistent. Yet they may also narrow assessment practice. Teachers become cautious. Innovative or negotiated assessment can appear risky. Dialogic feedback may be viewed as difficult to evidence. Holistic judgement may be suspected of subjectivity. Programme-level portfolios or longitudinal evidence may sit uneasily with systems designed around discrete module marks.

League tables and regulatory metrics intensify this pressure. When student satisfaction, completion, grade distribution, employability and external review outcomes become public indicators of institutional performance, assessment becomes not only pedagogical but reputational. Institutions may seek standardization not because it always improves learning, but because it appears safer. In such conditions, assessment reform can become conservative. Rubrics multiply. Feedback processes become procedural. Moderation becomes compliance. External examining becomes an audit ritual rather than a site of disciplinary dialogue.

The accountability–dialogue tension is therefore partly a clash of institutional logics. Dialogic assessment values interpretation, negotiation, development and contextual judgement. Quality assurance often values documentation, comparability, consistency and procedural evidence. The challenge is to show that these logics need not be enemies. Dialogue can be part of accountability when it is structured, documented and linked to defensible standards. Conversely, accountability becomes educationally weak when it treats standards as if they can be secured by documents alone.

2. Standards as Social and Dialogic

A central insight from Sadler, Bloxham and related assessment scholarship is that academic standards are not self-executing. They cannot be fully captured in criteria statements, rubrics or level descriptors. Standards are enacted through judgement. They are interpreted by assessors in relation to disciplinary expectations, task demands, student work and institutional contexts. This is why marker variability persists even when criteria are explicit.

The existence of variability does not mean assessors are careless. Studies of inter-tutor variability show that assessors often believe they are applying shared disciplinary standards, yet their tacit standards frameworks differ. One marker may weight argument more heavily; another may attend more to evidence; another may value originality, structure or methodological rigour. These differences are not always visible in written criteria. They emerge in the act of judging.

This creates a paradox. Quality assurance often assumes that reliability increases as standards are made more explicit in writing. But if standards are partly tacit and interpretive, writing them down is necessary but insufficient. More detailed rubrics may clarify some expectations, yet they cannot remove the need for judgement. Indeed, excessive rubric reductionism may create a false impression of objectivity while ignoring the holistic and disciplinary nature of academic quality.

The alternative is to recognize standards as social achievements. Standards are maintained through communities of practice: academics, external examiners, professional bodies and disciplinary networks who discuss, compare and refine judgement over time. Disciplined dialogue can contribute to reliability, but agreement must still be evaluated against the intended learning outcomes and evidence of performance. Assessors become more aligned not by silently reading criteria, but by examining student work together, explaining judgements, challenging interpretations and calibrating standards.

This view does not undermine standards. It makes their maintenance more honest. If standards are socially constructed, then institutions should make that construction visible and accountable. Students, staff and external stakeholders should understand that academic judgement involves professional interpretation, and that quality assurance depends on the strength of the interpretive community. This is not relativism. It is a more accurate account of how standards actually operate.

3. Moderation as Institutional Dialogue

Moderation is often treated as a technical quality procedure. Marks are checked, samples are reviewed, grade distributions are examined, and external examiners confirm that processes were followed. But moderation can be understood more deeply as institutional dialogue about standards. At its best, moderation is not merely a mechanism for detecting errors after marking. It is a professional learning process through which assessors build shared judgement before, during and after assessment.

Pre-marking calibration is especially important. Staff examine exemplars or sample scripts, apply criteria independently, compare judgements and discuss differences. This process surfaces tacit assumptions. Why did one assessor see the work as upper second-class while another saw it as lower second-class? Was the difference about evidence, structure, originality, disciplinary voice or interpretation of the criterion? Such discussion helps staff align their standards frameworks.

Social moderation extends this process. Instead of treating judgement as private and then checking it statistically, social moderation treats judgement as collectively developed. Assessors deliberate over borderline cases, compare examples across grade bands and refine their shared understanding of quality. This is dialogic assessment at the level of staff practice. The student may not be present, but the standards governing student judgement are being maintained through dialogue.

Comparative judgement offers another route. Pollitt's (2012) work on comparative judgement shows that assessors may be more reliable when

comparing two pieces of work than when assigning absolute marks against criteria. Adaptive comparative judgement operationalizes this insight by using repeated pairwise comparisons to generate rank orders. It offers a systematic use of holistic judgement, although reliability, shared biases and alignment with learning outcomes still require evaluation.

This does not mean comparative judgement should replace all assessment. Rather, it can support calibration, benchmarking and standards discussion. Staff might use comparative judgement to explore standards across modules, institutions or cohorts. Students might use comparative judgement to develop evaluative judgement. External examiners might engage with comparative samples to discuss threshold standards. The method offers one way of making holistic judgement more systematic without collapsing it into atomized criteria.

Moderation should therefore be reframed as a core site of assessment dialogue. Institutions should invest in it not only to satisfy audit requirements but to develop the professional judgement on which standards depend. Effective moderation requires time, trust, disciplinary expertise and leadership. It cannot be reduced to form completion.

4. External Examining and Quality Codes

External examining has long been a major mechanism for assuring standards, especially in the UK and systems influenced by it. External examiners are expected to provide independent oversight of academic standards, assessment processes and student achievement. In principle, this introduces external disciplinary judgement into institutional assessment. In practice, the system often relies heavily on sampling, documentation and retrospective review.

The limitation is that external examining may not sufficiently calibrate standards across individuals and institutions. If external examiners review samples after marking has occurred, they can identify anomalies or concerns, but they may not deeply align assessors' standards frameworks. If their role is mainly procedural, they may confirm that rules were

followed without substantially engaging in the social construction of standards. Bloxham and Price's (2015) critique is important here: external examining without meaningful calibration may provide weaker assurance than policy assumes.

A dialogic approach would reform external examining so that it becomes more actively developmental and standards-focused. External examiners could participate in pre-assessment calibration, cross-institutional standards workshops, comparative judgement exercises and programme-level review of assessment evidence. Their role would shift from retrospective auditor to member of a wider disciplinary standards community. This would strengthen rather than weaken accountability because it would address the interpretive basis of standards directly.

Quality codes would also need to accommodate more diverse forms of evidence. Traditional quality systems often assume discrete assessments, marks, criteria, samples and module-level decisions. Programmatic and dialogic assessment systems generate different evidence: portfolios, feedback action plans, mentor records, peer review logs, longitudinal progress data, calibration records, holistic committee decisions and student reflections on uptake. Regulators and quality agencies need frameworks that can judge the robustness of such systems without forcing them back into single-task measurement models.

Transcripts may also require reform. Conventional transcripts record modules, credits and grades. They say little about the developmental evidence underlying achievement, the capabilities demonstrated across time, or the student's evaluative judgement. A dialogic or programmatic approach might require richer records: achievement profiles, competency narratives, portfolio summaries or verified capability statements. Such reforms raise risks of complexity and comparability, but they also offer a more valid representation of learning than lists of isolated marks.

National quality agencies would need to adjust their expectations. They should ask not only whether standards are written down, but how communities calibrate them. Not only whether feedback was provided,

but whether students had opportunities to use it. Not only whether marks were moderated, but how assessors developed shared judgement. Not only whether assessment decisions are documented, but whether the documentation captures meaningful evidence of development.

5. Preserving Accountability Without Rubric Reductionism

The danger in quality assurance is that accountability becomes confused with standardization. Standardization can support fairness where tasks are simple or where consistency of procedure is essential. But in complex higher education assessment, excessive standardization may reduce validity. A dissertation, design project, clinical performance, reflective portfolio or critical essay cannot be judged well by mechanical application of isolated descriptors alone. Holistic professional judgement is necessary.

The aim should therefore be defensible judgement, not merely standardized procedure. Defensible judgement requires several conditions: clear standards, trained assessors, calibration, moderation, transparency, documentation, and opportunities for challenge or review. It also requires honesty about interpretation. Institutions should not pretend that marks are produced by neutral instruments. They should show how professional judgement is formed and checked.

This has implications for negotiated and dialogic assessment. Negotiation does not mean that standards are abandoned. It means that students may participate in interpreting criteria, choosing evidence, proposing topics or discussing task formats within boundaries. Quality assurance can accommodate this if the boundaries are explicit: fixed learning outcomes, documented criteria, approved assessment options, records of negotiation, and moderation of final judgements. The issue is not whether students had a voice, but whether the resulting assessment remains aligned, fair and transparent.

Programmatic assessment offers a parallel lesson. It may appear less standardized because it uses multiple data points and holistic decisions. Yet it can be highly accountable if it has clear outcomes, robust evidence

collection, trained mentors, independent committee review, documented decision rules and external calibration. A single standardized exam may look more reliable but capture less of the relevant competence. Quality assurance should therefore consider validity across time, not only consistency at one point.

5.4 Generative AI and the Future of Assessment Dialogue

Generative artificial intelligence has changed the assessment problem in higher education. ChatGPT and its successors did not create the weaknesses of traditional assessment, but they exposed them with unusual force. Product-based, unsupervised and text-heavy assessments had long rested on fragile assumptions: that the submitted product was authored by the student, that performance on a single task could validly represent learning, and that written output could be treated as a stable proxy for understanding. Generative AI destabilizes all three assumptions. It can produce fluent essays, summaries, reports, reflections, code, plans and explanations. It can imitate the surface features of academic work. It can help students brainstorm, draft, revise and polish. In doing so, it forces higher education to confront a question that this book has asked throughout: what exactly are we assessing when we assess?

The immediate response to generative AI has often been anxiety about cheating. This anxiety is understandable. If students can outsource the production of an essay, the validity of the assessment inference is threatened. A submitted product may no longer show what the student knows, can do or can judge. Yet the deeper disruption is not only academic misconduct. Generative AI reveals the limits of the measurement paradigm. If assessment depends primarily on isolated products produced outside view, then it is vulnerable to any technology that can simulate such products. The problem is not merely that AI makes cheating easier. The problem is that many assessment systems were already

too dependent on products detached from process, judgement and dialogue.

This section argues that generative AI strengthens the case for dialogic assessment. The durable response to AI is not an endless arms race of detection, surveillance and prohibition. It is the redesign of assessment around human capabilities that must be demonstrated, defended and developed: situated judgement, reasoning, ethical decision-making, disciplinary interpretation, critical appraisal, responsiveness to feedback and the capacity to explain one's process. These capabilities are not secured by hiding from AI. They are cultivated through assessment designs that foreground dialogue, evidence over time, and evaluative judgement.

1. Disruption: AI and the Fragility of Product-Based Assessment

Generative AI has particularly disrupted product-based, unsupervised assessments. Essays, take-home exams, reflective commentaries, business reports, literature reviews and short-answer tasks are all affected because AI can generate plausible versions of these products. This does not mean such tasks are obsolete. It means they can no longer be treated as simple evidence of unaided student capability unless the assessment design explains what kind of AI use is permitted, how process is evidenced, and how authorship or understanding is verified.

The disruption is epistemic as much as procedural. In traditional assessment, the product often stands in for the learner. The essay is marked as though it transparently represents the student's knowledge. Generative AI breaks this transparency. The product may represent the student's understanding, the AI's output, the student's editing of AI output, a peer's contribution, or some combination. The submitted artefact becomes ambiguous evidence.

This ambiguity exposes a weakness that existed before AI. Students have always received help: from tutors, friends, editing services, model answers, online resources and institutional privilege. Generative AI

democratizes some forms of assistance while also introducing new inequities and risks. The assessment problem is therefore not simply to restore a lost purity of authorship. It is to design assessments where the relationship between student capability, assistance and submitted evidence is explicit, interpretable and educationally meaningful.

Generative AI also accelerates the need to assess process. If the final text can be generated, then assessment must ask how the student formulated the problem, selected sources, evaluated outputs, revised arguments, justified choices and responded to critique. This is precisely where dialogic assessment becomes relevant. Dialogue makes the student's reasoning visible. It asks students not only to submit an answer but to explain, defend, revise and situate it.

2. Dialogue as Response, Not Detection

The detection response seeks to identify AI-generated text and punish unauthorized use. Detection tools may have a place in institutional integrity systems, but they cannot carry the burden of assessment reform. Weber-Wulff et al. (2023) found substantial accuracy and reliability limitations in the tools they tested. Detection results therefore require contextual validation and should not alone establish misconduct. A detection-led approach can also damage trust, especially when students are treated primarily as suspects. A surveillance-heavy response intensifies the disciplinary logic of assessment without addressing the learning problem.

A dialogic response begins differently. It asks: what human capability should this assessment develop and evidence? Can the student explain the reasoning behind the work? Can they justify methodological choices? Can they critique an AI-generated answer? Can they respond to questions? Can they connect the work to a local context, professional dilemma or disciplinary debate? Can they show how feedback changed the product? Can they identify limitations and ethical implications? These questions move assessment away from mere product verification toward judgement and defence of understanding.

This does not mean abandoning assessment security. Dawson's (2021) work on assessment security is important because it reminds us that trust in credentials matters. Some assessment moments must be secure, especially where professional competence, safety or certification is at stake. But security should not be reduced to detection. Supervised performance, oral defence, viva-style questioning and interactive orals can strengthen assurance when access to assistance is appropriately controlled. Staged submissions, version histories and authentic tasks can supply supporting evidence, but do not by themselves establish unaided capability. The goal is to make the assessment inference trustworthy, not merely to catch misconduct after the fact.

Authentic assessment is often proposed as a solution to AI. The term must be used carefully. Some authentic tasks can still be supported by AI. A workplace report, policy brief or professional email may be highly AI-assisted in real life. The issue is not whether AI can help with the task, but whether the assessment requires students to exercise judgement about using AI well. Future-authentic assessment should prepare students for AI-integrated contexts, not pretend AI will disappear.

Interactive oral assessment and viva-style dialogue are especially important in this context. They allow students to explain reasoning in real time, respond to probing questions, connect work to experience and demonstrate ownership of ideas. Their value lies not in nostalgia for oral exams, but in making understanding responsive. A student may submit an AI-assisted product, but the dialogue can ask: Why did you choose this approach? What alternative did you reject? Which claim is weakest? How did you verify the evidence? What would you change if the context shifted? These questions assess judgement. Preparation, consistent questioning and accessible alternatives are needed so that oral fluency, anxiety or disability do not obscure the capability being assessed.

3. Programmatic and Two-Lane Approaches

No single assessment form can solve the AI problem. The most robust response is programmatic. A programme should gather multiple kinds of evidence across time: written products, drafts, peer feedback, oral explanations, reflective accounts, supervised tasks, practical demonstrations, AI-use logs, portfolios, mentoring conversations and high-stakes validation points. The logic resembles the programmatic assessment model discussed in Chapter 3: many low-stakes data points support learning and feedback, while a smaller number of high-stakes decisions are made from richer evidence.

In an AI era, the two-lane approach described by Bridgeman, Liu and Weeks (2024) offers a useful starting point. Adapted here for dialogic assessment, the open learning lane (lane 2) is developmental and dialogic. It includes low-stakes and medium-stakes tasks where students may use AI under clear conditions, experiment, receive feedback, critique outputs, revise and reflect. This lane is designed for learning. It assumes that students need guided practice in AI-rich environments. It includes peer review, feedback request sheets, AI-output critique, draft conferences, portfolios and reflective commentaries.

The secure lane (lane 1) is for validation. It includes fewer but more secure, dialogue-rich high-stakes points where students must demonstrate ownership, understanding and judgement. These may include interactive orals, vivas, supervised performances, oral defences of written work, capstone presentations, practical demonstrations or structured interviews. This lane is designed for assurance. It does not need to verify every sentence of every submitted product. It verifies that the student can reason with, defend and apply the capabilities the programme claims to assess.

The two lanes should be connected. Developmental tasks prepare students for validation tasks. AI-supported drafting can lead to oral defence. A portfolio of AI interactions can lead to a viva on evaluative judgement. Peer and AI feedback can feed a final reflective account. A

research proposal developed with AI assistance can be tested through questions about methodology, ethics and source credibility. The aim is not to separate learning from security, but to place security at carefully chosen points within a broader developmental architecture.

This model also reduces assessment panic. If every task is treated as a high-stakes proof of unaided authorship, institutions will default to surveillance and prohibition. If every task permits unlimited AI use without validation, standards may be weakened. A two-lane model recognizes that not all assessment moments have the same purpose. Some are for practice; some are for evidence; some are for certification. Dialogic assessment helps connect them.

4. AI as a Dialogic Partner: Potential and Risk

Generative AI can also function as a dialogic partner for low-stakes learning. Students can ask it to explain concepts, generate practice questions, critique drafts, simulate interviews, produce counterarguments, simplify complex readings, role-play clients or provide formative feedback. Used well, AI can increase access to practice dialogue. A student who cannot meet a tutor immediately can still receive prompts, questions and alternative explanations. AI can support rehearsal before human dialogue.

AI-generated feedback may also help address scale. In large cohorts, students often receive too little formative feedback before high-stakes submission. AI tools can provide rapid responses to drafts, highlight possible gaps, suggest questions and prompt reflection. Teachers may use AI to generate feedback drafts, identify common issues or create exemplars. These uses can extend assessment dialogue if they are embedded in human judgement and pedagogical design. Institutions should protect student data, ensure equitable access and require human review of consequential feedback (Miao & Holmes, 2023).

But the risks are serious. First, AI feedback may be inaccurate, generic or overconfident. It may reward surface fluency and miss disciplinary

depth. It may invent sources, misread context or reproduce dominant norms. Second, students may over-trust AI because it sounds authoritative. Novices are particularly vulnerable because they lack the evaluative judgement needed to detect weak feedback. Third, AI systems may reproduce linguistic, cultural, gendered, racial or disciplinary biases present in training data. Fourth, reliance on AI feedback may deskill educators if institutions use it to reduce human engagement rather than enhance it.

The most serious risk, from the perspective of this book, is the loss of recognition. Dialogic assessment is not only information exchange. It is a relational encounter in which the student is recognized as a developing knower. AI can simulate responsiveness, but it does not know the student in the ethical sense. It cannot offer genuine care, disciplinary membership, professional responsibility or human recognition. It can generate comments, but it cannot replace the relational dimension through which critique becomes trusted and identity-shaping.

Therefore, AI should not be treated as a substitute assessor or relational mentor. It is better understood as a low-stakes interlocutor, comparator and prompt generator. It can extend opportunities for practice, but human educators must remain responsible for standards, recognition, ethical judgement and high-stakes decisions. The danger is not that AI will participate in feedback. The danger is that institutions will use AI to hollow out human dialogue while claiming to scale it.

5. Evaluative Judgement About AI

The most important educational response to generative AI is to make critical appraisal of AI outputs a core form of evaluative judgement. Students now need to judge not only their own work and peers' work, but also machine-generated work. They must ask: Is this answer accurate? What assumptions does it make? What evidence is missing? What bias might it contain? Does it meet disciplinary standards? What has been hallucinated? What is persuasive but shallow? What ethical risks arise if this output is used?

This extends the concept of evaluative judgement developed throughout the book. In the pre-AI assessment world, students needed to judge quality in human-produced texts, performances and professional actions. In the AI era, they must judge hybrid human-machine outputs and processes. The capability is no longer optional. Graduates will encounter AI-generated drafts, summaries, analyses, recommendations and decisions in workplaces and public life. They must be able to interrogate them.

Dialogic assessment is well suited to cultivating this capability. Students can compare AI-generated answers with human answers, peer answers, expert exemplars and disciplinary criteria. They can revise AI outputs and justify changes. They can debate whether an AI-generated argument is valid. They can identify missing perspectives. They can document prompts and evaluate the process. They can defend when AI use was appropriate and when it was not. These tasks make AI a site for judgement rather than a shortcut around judgement.

For example, in a sociology module, students might ask AI to produce an explanation of social capital, then critique the output using Bourdieu (1986) and Coleman (1988). In a literature module, students might compare an AI interpretation of a poem with scholarly readings. In a professional ethics module, students might evaluate an AI-generated policy recommendation for bias, feasibility and accountability. In each

case, the assessment is not simply “use AI” or “do not use AI.” It is “show that you can judge AI.”

This shifts academic integrity from prohibition to capability. Students must learn what responsible AI use means in a discipline. They must understand attribution, verification, privacy, bias, intellectual dependence and the difference between assistance and substitution. They must be able to explain their choices. In this sense, AI literacy becomes part of assessment literacy.

5.5 Epilogue: Assessment as a Practice of Recognition

The argument of this book has been built around a simple but often obscured claim: assessment has always been a relationship between people. Even when it appears to be a technical act, even when mediated by rubrics, grade descriptors, moderation forms, marking schemes, algorithms or quality assurance codes, assessment remains an encounter in which one person’s work, judgement, effort and emerging identity are interpreted by others. The measurement paradigm disguised this relationship by presenting assessment as the grading of objects. A script, essay, portfolio, exam answer or project appeared to be the thing being measured. The student became backgrounded; the assessor became an instrument; the grade became the truth.

This book has tried to make the relationship visible again. It has argued that assessment is not only a matter of accuracy, reliability or comparability, though these remain important. It is also a matter of trust, recognition, interpretation, power, dialogue and development. To assess is not merely to assign value to a finished product. It is to participate in the formation of a learner’s understanding of quality, capability and self. Assessment tells students not only what their work is worth, but what kind of knower they are becoming and whether they belong within a community of judgement.

The central contribution of the book is the Dialogic Assessment Framework. The DAF proposes that many of the deepest problems of assessment are not solved by better rubrics alone. Disengagement persists when feedback has no place to go. Inequity persists when standards remain hidden beneath technical language. Distrust persists when judgement is experienced as distant, unexplained or punitive. Students continue to game criteria when assessment rewards compliance more than understanding. Generative AI exposes the fragility of assessment systems built on unsupervised products and weak evidence of judgement. These problems require not only clearer instruments but better conversations: conversations about standards, evidence, feedback, power, identity, process and future action.

This does not mean that every assessment problem can be solved by talking more. Dialogue, as this book has insisted, is not a decorative technique. It is not a tutorial added after grading, a peer review sheet filled in hurriedly, or a token invitation to comment on criteria that have already been decided. Dialogue is a disciplined educational relation. It requires structure, trust, continuity, access to standards, opportunities for uptake and ethical attention to voice. The DAF has therefore distinguished surface dialogic features from deep dialogic conditions. Without those conditions, dialogic assessment becomes dialogue-washing: progressive in appearance but unchanged in substance.

The ethical foundation of this argument reaches beyond assessment technique. Honneth's (1995) theory of recognition helps us see that students are formed through the ways institutions recognize or misrecognize them. Assessment is one of the most consequential forms of educational recognition. A student who receives serious, demanding, respectful feedback is being addressed as someone capable of growth. A student whose work is dismissed, reduced to a number, or judged through opaque standards may experience assessment as misrecognition. Recognition does not mean praise without standards. It means that

judgement is offered in a way that acknowledges the student as a developing knower.

Bakhtin's (1993) notion of answerability deepens this ethical claim. Feedback and grading are not neutral outputs. They are utterances addressed to someone. They enter a student's world, invite or foreclose response, and shape future action. The assessor is therefore answerable for the form of judgement, not because they can control every effect, but because assessment is an ethical act of address. A comment that cannot be questioned, a grade that cannot be understood, or a standard that cannot be discussed closes dialogue at the very moment when learning requires response.

Buber and Freire extend the argument further. Buber's (1970) distinction between I–It and I–Thou illuminates the danger of assessment systems that turn students into objects of measurement. When students are treated only as scripts, scores or data points, the assessment relation becomes I–It. A dialogic approach seeks, however imperfectly, to restore the I–Thou dimension: the recognition that the learner is a person in formation, not merely a performance to be classified. Freire (1970) reminds us that dialogue is also political. Assessment can domesticate students into compliance, or it can help them become critical co-investigators of knowledge, standards and action. The ethical question is therefore not only whether assessment is fair in procedure, but whether it enlarges or diminishes students' agency.

This leads to a necessary assessment humility. Higher education has often spoken as though standards were fully explicit, objective and stable. Yet the evidence reviewed throughout this book suggests otherwise. Standards are partly tacit. They are socially constructed within communities of practice. They are interpreted through disciplinary histories, institutional cultures, assessor experience and professional judgement. They are contested, negotiated and revised. The pretence of perfectly objective measurement is not only false; it can be harmful. It hides the human judgement on which assessment depends and makes

students responsible for decoding standards that have never been fully disclosed.

Dialogue is an honest alternative. It does not deny standards. It refuses to pretend that standards become fair simply because they are written down. It recognizes that standards are made learnable through exemplars, comparison, moderation, feedback, peer discussion, mentoring and opportunities to act. It also recognizes that reliability itself is often produced through dialogue among assessors: calibration, social moderation, shared exemplars and professional deliberation. Dialogue therefore does not weaken rigour. It makes rigour more visible, more accountable and more educative.

The future of assessment depends on this honesty. In an age of mass higher education, international mobility, widening participation and generative AI, the old settlement is breaking down. Students are more diverse. Knowledge practices are more complex. Technologies can produce convincing academic products without guaranteeing understanding. Employers and societies need graduates who can judge evidence, critique outputs, explain reasoning, work with others and continue learning. Assessment systems that merely collect finished products and assign grades are poorly suited to this world.

A dialogic vision of assessment offers another path. It imagines assessment not as the terminal judgement of a finished product, but as a continuing conversation through which learners become capable of judging well for themselves. This conversation begins with external standards, but it does not end there. Students encounter criteria, exemplars, teacher feedback, peer review, programme evidence and professional expectations. They compare, question, revise, defend and reflect. Over time, external judgement becomes internal evaluative capacity. The voice of the teacher, the peer, the mentor and the discipline becomes part of the student's own capacity to ask: Is this good enough? How do I know? What evidence supports my judgement? What must change? What responsibility follows?

This is the truest aim of university education. Universities should not produce graduates who merely wait to be judged. They should produce graduates who can participate in judgement: of knowledge, evidence, practice, technology, institutions and themselves. Such judgement is not solitary. It is learned through relationships with communities and traditions, and through dialogue. But its mature form is sustainable. The graduate carries forward the capacity to evaluate, to seek feedback, to revise understanding and to act responsibly when no teacher is present.

The book therefore closes with a claim that is both intellectual and ethical. Assessment is never only measurement. It is a relationship of judgement that can either conceal or reveal its human character. When concealed, assessment becomes a technology of classification: efficient, defensible perhaps, but often alienating and inequitable. When revealed and redesigned as dialogue, assessment can become a practice of recognition, development and shared responsibility.

Beyond the rubric lies not the abandonment of standards, but their humanization. Beyond grading lies not softness, but a more demanding question: how do we help students understand, inhabit, question and eventually carry standards for themselves? Beyond assessment as control lies assessment as conversation. That conversation is unfinished, as all genuine dialogue must be. But it is precisely in its unfinishedness that assessment becomes educational: a continuing exchange through which students, teachers and institutions learn what it means to judge well.

References

- Alexander, R. J. (2008). *Towards dialogic teaching: Rethinking classroom talk* (4th ed.). Dialogos.
- Anderson, G., Boud, D., & Sampson, J. (1996). *Learning contracts: A practical guide*. Routledge.
- Bakhtin, M. M. (1981). *The dialogic imagination: Four essays* (M. Holquist, Ed.; C. Emerson & M. Holquist, Trans.). University of Texas Press.
- Bakhtin, M. M. (1986). *Speech genres and other late essays* (C. Emerson & M. Holquist, Eds.; V. W. McGee, Trans.). University of Texas Press.
- Bakhtin, M. M. (1993). *Toward a philosophy of the act* (V. Liapunov, Trans.; V. Liapunov & M. Holquist, Eds.). University of Texas Press.
- Baume, D., & Yorke, M. (2002). The reliability of assessment by portfolio on a course to develop and accredit teachers in higher education. *Studies in Higher Education*, 27(1), 7–25. <https://doi.org/10.1080/03075070120099340>
- Beaumont, C., O'Doherty, M., & Shannon, L. (2011). Reconceptualising assessment feedback: A key to improving student learning? *Studies in Higher Education*, 36(6), 671–687. <https://doi.org/10.1080/03075071003731135>
- 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>
- Bloxham, S., & Price, M. (2015). External examining: Fit for purpose? *Studies in Higher Education*, 40(2), 195–211. <https://doi.org/10.1080/03075079.2013.823931>
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151–167. <https://doi.org/10.1080/713695728>

- Boud, D., & Dawson, P. (2023). What feedback literate teachers do: An empirically-derived competency framework. *Assessment & Evaluation in Higher Education*, 48(2), 158–171. <https://doi.org/10.1080/02602938.2021.1910928>
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698–712. <https://doi.org/10.1080/02602938.2012.691462>
- Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Greenwood Press.
- Bridgeman, A., Liu, D., & Weeks, R. (2024). *Program level assessment design and the two-lane approach*. Teaching@Sydney, University of Sydney. <https://educational-innovation.sydney.edu.au/teaching%40sydney/program-level-assessment-two-lane/>
- Buber, M. (1970). *I and Thou* (W. Kaufmann, Trans.). Charles Scribner's Sons.
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219–233. <https://doi.org/10.1080/03075070600572132>
- 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>
- Carless, D. (2009). Trust, distrust and their impact on assessment reform. *Assessment & Evaluation in Higher Education*, 34(1), 79–89. <https://doi.org/10.1080/02602930801895786>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Carless, D., & Winstone, N. (2023). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, 28(1), 150–163. <https://doi.org/10.1080/13562517.2020.1782372>

- CAST. (2024). *CAST Universal Design for Learning Guidelines version 3.0*.
<https://udlguidelines.cast.org/>
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94(Supplement), S95–S120. <https://doi.org/10.1086/228943>
- Cook-Sather, A., Bovill, C., & Felten, P. (2014). *Engaging students as partners in learning and teaching: A guide for faculty*. Jossey-Bass.
- Dawson, P. (2021). *Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education*. Routledge.
- Foucault, M. (1977). *Discipline and punish: The birth of the prison* (A. Sheridan, Trans.). Allen Lane.
- Freire, P. (1970). *Pedagogy of the oppressed* (M. B. Ramos, Trans.). Herder and Herder.
- Gergen, K. J. (2009). *Relational being: Beyond self and community*. Oxford University Press.
- Handley, K., & Williams, L. (2011). From copying to learning: Using exemplars to engage students with assessment criteria and feedback. *Assessment & Evaluation in Higher Education*, 36(1), 95–108. <https://doi.org/10.1080/02602930903201669>
- Healey, M., Flint, A., & Harrington, K. (2014). *Engagement through partnership: Students as partners in learning and teaching in higher education*. Higher Education Academy. <https://repository.londonmet.ac.uk/5176/>
- Hill, J., & West, H. (2022). Dialogic feed-forward in assessment: Pivotal to learning but not unproblematic. *Teaching & Learning Inquiry*, 10. <https://doi.org/10.20343/teachlearninqu.10.20>
- Honneth, A. (1995). *The struggle for recognition: The moral grammar of social conflicts* (J. Anderson, Trans.). Polity Press.
- Hughes, G. (2014). *Ipsative assessment: Motivation through marking progress*. Palgrave Macmillan. <https://doi.org/10.1057/9781137267221>

- Jessop, T., & Tomas, C. (2017). The implications of programme assessment patterns for student learning. *Assessment & Evaluation in Higher Education*, 42(6), 990–999. <https://doi.org/10.1080/02602938.2016.1217501>
- Klenowski, V. (2002). *Developing portfolios for learning and assessment: Processes and principles*. Routledge. <https://doi.org/10.4324/9780203825921>
- Li, H., Xiong, Y., Zang, X., Kornhaber, M. L., Lyu, Y., Chung, K. S., & Suen, H. K. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education*, 41(2), 245–264. <https://doi.org/10.1080/02602938.2014.999746>
- Littleton, K., & Mercer, N. (2013). *Interthinking: Putting talk to work*. Routledge.
- Loosveld, L. M., van Gerven, P. W. M., Vanassche, E., & Driessen, E. W. (2020). Mentors' beliefs about their roles in health care education: A qualitative study of mentors' personal interpretative framework. *Academic Medicine*, 95(10), 1600–1606. <https://doi.org/10.1097/ACM.00000000000003159>
- Mentzer, N., Lee, W., & Bartholomew, S. R. (2021). Examining the validity of adaptive comparative judgment for peer evaluation in a design thinking course. *Frontiers in Education*, 6, 772832. <https://doi.org/10.3389/feduc.2021.772832>
- Mercer, N. (2000). *Words and minds: How we use language to think together*. Routledge.
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>
- Molloy, E., & Bearman, M. (2019). Embracing the tension between vulnerability and credibility: 'Intellectual candour' in health professions education. *Medical Education*, 53(1), 32–41. <https://doi.org/10.1111/medu.13649>
- Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, 46(5), 756–778. <https://doi.org/10.1080/02602938.2020.1823314>

- 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>
- Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102–122. <https://doi.org/10.1080/02602938.2013.795518>
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74–98. <https://doi.org/10.1016/j.edurev.2017.08.004>
- Pitt, E., & Norton, L. (2017). ‘Now that’s the feedback I want!’ Students’ reactions to feedback on graded work and what they do with it. *Assessment & Evaluation in Higher Education*, 42(4), 499–516. <https://doi.org/10.1080/02602938.2016.1142500>
- Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 19(3), 281–300. <https://doi.org/10.1080/0969594X.2012.665354>
- Power, M. (1997). *The audit society: Rituals of verification*. Oxford University Press.
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435–448. <https://doi.org/10.1080/02602930902862859>
- Rust, C., O’Donovan, B., & Price, M. (2005). A social constructivist assessment process model: How the research literature shows us this could be best practice. *Assessment & Evaluation in Higher Education*, 30(3), 231–240. <https://doi.org/10.1080/02602930500063819>

- Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535–550. <https://doi.org/10.1080/02602930903541015>
- Schut, S., Driessen, E., van Tartwijk, J., van der Vleuten, C., & Heeneman, S. (2018). Stakes in the eye of the beholder: An international study of learners' perceptions within programmatic assessment. *Medical Education*, 52(6), 654–663. <https://doi.org/10.1111/medu.13532>
- Scoggins, J., & Winter, R. (1999). The patchwork text: A coursework format for education as critical understanding. *Teaching in Higher Education*, 4(4), 485–499. <https://doi.org/10.1080/1356251990040405>
- Shulman, L. S. (2005). Signature pedagogies in the professions. *Daedalus*, 134(3), 52–59. <https://www.amacad.org/publication/daedalus/signature-pedagogies-professions>
- Strathern, M. (Ed.). (2000). *Audit cultures: Anthropological studies in accountability, ethics and the academy*. Routledge.
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467–481. <https://doi.org/10.1007/s10734-017-0220-3>
- To, J., & Carless, D. (2016). Making productive use of exemplars: Peer discussion and teacher guidance for positive transfer of strategies. *Journal of Further and Higher Education*, 40(6), 746–764. <https://doi.org/10.1080/0309877X.2015.1014317>
- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice*, 14(3), 281–294. <https://doi.org/10.1080/09695940701591867>

- van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, 34(3), 205–214. <https://doi.org/10.3109/0142159X.2012.652239>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19, Article 26. <https://doi.org/10.1007/s40979-023-00146-z>
- Wegerif, R. (2013). *Dialogic: Education for the Internet Age*. Routledge.
- Wertsch, J. V. (1998). *Mind as action*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195117530.001.0001>
- Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education*, 18(3), 285–297. <https://doi.org/10.1080/13562517.2012.719154>
- Yosso, T. J. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. *Race Ethnicity and Education*, 8(1), 69–91. <https://doi.org/10.1080/1361332052000341006>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2