

RECALIBRATING LANGUAGE ASSESSMENT: TOWARD A METACOGNITIVE PARADIGM FOR 21ST-CENTURY ESPAK Karunarathne¹ and WMPYB Rathnayake²**Abstract**

This narrative literature review critically examines the evolving role of language assessment in English for Specific Purposes (ESP), with reference to engineering education in South Asia. Historically shaped by structuralist and summative paradigms, assessment has often constrained learner agency and metacognitive growth, two capacities that are essential for academic success and professional readiness in STEM. Using an integrative review approach, the study synthesizes seminal theories and contemporary empirical research to map the shift from static, test-oriented practices to dynamic, formative, and context-sensitive models. The methodology was guided by a purposive sampling strategy, selecting seven milestone texts for their theoretical influence and historical significance, complemented by recent empirical studies addressing dynamic assessment, integrated skill testing, and AI-supported feedback. This dual focus was reinforced through a Global South-aware approach, which prioritized open-access peer-reviewed literature, cross-referenced seminal works, and drew on targeted searches in widely accessible databases such as Google Scholar, Scopus (via partial institutional access), and ERIC—an adaptation made necessary by the restricted access to subscription resources in Sri Lanka and comparable contexts. Findings indicate that integrative tasks, reflective journaling, peer feedback, and digitally mediated scaffolds promote self-regulation, critical thinking, and adaptive language use in multilingual engineering environments. Ethical imperatives, including dismantling native-speaker norms, introducing intelligibility-focused rubrics, and strengthening teacher assessment literacy, are also emphasized. The review argues that ESP assessment must be re-engineered beyond its gatekeeping role toward an ethically grounded, metacognitively enriched paradigm. In such a framework, assessment functions not only as a measure of competence but as a catalyst for learner autonomy, disciplinary readiness, and educational equity. By positioning assessment as a pedagogical catalyst for transformation, this study contributes a regionally grounded yet globally relevant perspective for reshaping 21st-century ESP practices.

Keywords: Assessment as Learning, Dynamic Assessment, ESP Curriculum, Metacognition, Self-Regulated Learning

¹Department of Languages, University of Moratuwa, Sri Lanka.

Email: ashminik@uom.lk



<https://orcid.org/0009-0003-5202-8152>

²Department of Languages, University of Moratuwa, Sri Lanka.

Email: prasannar@uom.lk



<https://orcid.org/0000-0003-2835-3249>



Proceeding of the 3rd Desk Research Conference – DRC 2025 © 2025 by The Library, University of Kelaniya, Sri Lanka is licensed under [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)

Received date: 08.05.2025

Accepted date: 02.07.2025

Print Publishing Date: 31.10.2025

Web Publishing Date: 31.10.2025

Introduction

Over the past century, language testing has evolved from a mechanistic process of quantifying linguistic items to a mature, multi-faceted practice embedded in pedagogical, social, and ideological contexts. Traditionally fortified by structural linguistics and behaviorist learning theories, first-generation models of assessment were preoccupied with discrete-point testing, reliability concerns, and psychometric precision (Davies, 2013; Rahman, 2017). Growing dissatisfaction with the limited communicative validity of such tests, however, led to a paradigmatic shift toward holistic assessments motivated by sociolinguistic, constructivist, and sociocultural perspectives (Douglas & Abeywickrama, 2018; Fulcher & Harding, 2022).

This transition holds particular importance for English for Specific Purposes (ESP), where assessment must not only indicate general language proficiency but also measure learners' ability to perform discipline-specific communicative tasks. In STEM fields such as engineering—where language is inscribed in technical composition, data representation, and collaborative work—assessment becomes central to shaping students' academic and professional identities. As Çelik, Çelik, and Coombe (2021) argue, ESP learners must engage not only with linguistic structures but also with the rhetorical and cognitive demands of their domain; thus, assessment becomes a site of tactical and situated sense-making.

Despite such theoretical advances, assessment in many ESP contexts remains dominated by summative, grammar-based approaches that neglect the cognitive, reflective, and ethical dimensions of learning. In South Asian high stakes testing cultures such as Sri Lanka, certification is still valued over capacity development, with limited pedagogical impact and weak alignment to learners' communicative needs (Rahman et al., 2018; Islam et al., 2021). Such practices inhibit students' metacognitive growth—particularly their ability to plan, monitor, and regulate their learning—thereby undermining the broader goal of fostering independent, critically aware professionals.

Accordingly, this review seeks to reorient ESP assessment in engineering education through a metacognitive pedagogical lens. It argues that assessment must move beyond static measurement of language output toward a dynamic, formative, and ethically defensible process that cultivates student agency and disciplinary readiness. Metacognition, defined as individuals' awareness and regulation of cognition (Flavell, 1979), has become a powerful determinant of academic achievement, particularly in self-regulated learning environments (Panadero, 2017). Embedding metacognitive goals into assessment can transform the classroom into a reflective, judgment-free space that fosters strategic adaptation and professional growth.

Drawing on a heterogeneous body of secondary sources to trace historical evolution, thematic trends, and contested issues in language testing, this review adopts a narrative literature review method. It pays particular attention to how ESP assessment has responded to theoretical innovations such as Dynamic Assessment (Vygotsky, 1978; Poehner, 2008), task-based integrative design, critical pedagogy (Freire, 1970), and Assessment as Learning (Salaberry & Weideman, 2024). The review synthesizes insights from global and South Asian scholarship to offer a comprehensive and context-sensitive account of the ESP assessment landscape.

The broader objectives of this review are threefold:

1. To trace the historical and theoretical evolution of language testing within ESP pedagogy;
2. To critically examine how current assessment practices enable or hinder the development of metacognitive skills; and
3. To propose pedagogical and policy recommendations for re-engineering assessment practices as catalysts for transformative learner development in ESP–Engineering programs.

By interlacing diverse theoretical perspectives and empirical findings, this paper contributes to ongoing debates on assessment reform, particularly within underrepresented and linguistically marginalized educational environments. It emphasizes the urgent need to reconceptualize assessment not merely as a tool for evaluating what students know but as a process for shaping how they think, act, and learn as future professionals.

This review is guided by the overarching question: How have language assessment paradigms evolved to support metacognitive development within ESP–Engineering education, particularly in postcolonial multilingual contexts such as South Asia?

Methodology / Research Design

This study adopts a narrative literature review approach to critically examine the evolution of language testing and its potential to foster metacognitive development in English for Specific Purposes (ESP) within engineering

education. In line with the requirements of desk research, the study relies exclusively on secondary sources and emphasizes a critical synthesis of theoretical and empirical works.

Following Whittemore and Knafel's (2005) integrative review method, which accommodates both theoretical debates and empirical studies, the review integrates perspectives from applied linguistics, pedagogy, psychology, and educational ethics to illuminate interdisciplinary insights. As the study originates from a Global South context, it also adopts the methodological orientation proposed by Rathnayake (2024), who demonstrates how a Global South–critical approach can be employed to overcome limited access to subscription-based databases that constrain scholarly work in Sri Lanka and similar settings. Specifically, the methodology draws on:

- prioritizing open-access peer-reviewed publications and authoritative handbooks,
- consulting recent encyclopedic entries and critical handbooks (e.g., Fulcher & Harding, 2022),
- extensive cross-referencing of seminal citations, and
- targeted keyword searches in widely accessible platforms such as Google Scholar, Scopus (via partial institutional access), and ERIC.

A purposive sampling strategy was employed to ensure that the included literature was (a) theoretically foundational, (b) methodologically rigorous, and (c) contextually relevant to ESP and metacognition. Seven milestone texts (e.g., Davies, 2013; Fulcher & Harding, 2022; Douglas & Abeywickrama, 2018) were selected for their enduring influence in shaping global debates on assessment theory and practice. These were complemented by recent empirical studies (e.g., Huang & Jiang, 2020; Shrestha, 2020; Islam et al., 2021) that examine alternative assessment practices, dynamic assessment, and AI-supported feedback systems in both global and South Asian ESP contexts. This dual focus connects canonical works that provide theoretical and historical baselines with emerging evidence that addresses contemporary challenges in multilingual, resource-constrained educational systems.

The search strategy involved keywords such as *ESP assessment*, *metacognition in language testing*, *dynamic assessment in engineering*, and *alternative assessment in STEM*. Inclusion criteria limited sources to peer-reviewed works published between 2000 and 2024, with priority given to studies addressing either ESP pedagogy or metacognitive dimensions of assessment in higher education.

The collected studies were analyzed through thematic coding, beginning with broad categories such as historical development, assessment paradigms, metacognitive approaches, skill integration, and ethical concerns. Through iterative reading and refinement, these categories were consolidated into three interpretive dimensions:

1. the theory and history of language assessment,
2. the role of metacognition in learner agency and self-regulation, and
3. the ethical and contextual reinterpretation of evaluation in ESP–Engineering.

In adopting Rathnayake's (2024) Global South–critical stance, this study acknowledges the inequities in access to paywalled knowledge yet demonstrates methodological rigor by triangulating open-access, seminal, and locally relevant scholarship. This ensures that the review remains both globally informed and regionally grounded.

While narrative reviews are inherently limited by subjectivity and replicability issues, this study mitigates such concerns through transparency in selection criteria, explicit rationale for text inclusion, and alignment with critical regional contexts. This reflexive design enables a context-sensitive exploration of how ESP assessment may evolve from certification-driven practice toward metacognitively enriched, developmental pedagogy.

Results and Discussion

Historical Trajectory of Language Assessment

The early stages of language testing were governed by structuralist and behaviorist philosophies, viewing language as a collection of discrete units that were best tested through isolated grammar and vocabulary tests (Davies, 2013). As illustrated in Fig. 1 the trajectory was a dynamic shift from a structuralist to a metacognitive orientation. The psychometric practices emphasized objectivity and replicability, leading to the application of multiple-choice and fill-in-the-blank exercises yielding quantifiable but pedagogically limited information. Rahman (2017) suggests, such paradigms were a positivist epistemology, with statistical validity over communicative competence.

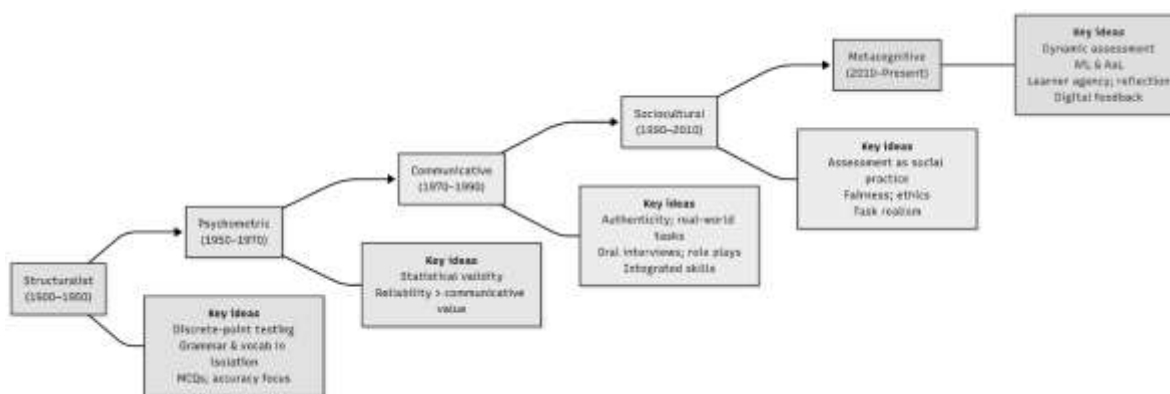


Figure 01: Progression of Language Assessment Paradigms (1900-2020).

Note. Created By Authors.

The trend shifted toward communicative language testing by the 1970s and 1980s, with the influence of Hymes' (1972) notion of communicative competence and Canale and Swain's (1980) integrative model. Testing began to reflect real-world tasks such as oral interviews, role plays, integrated skill tests, with a purposeful focus on authenticity, contextual relevance, and learner agency (Douglas & Abeywickrama, 2018; Fulcher & Harding, 2022). But this shift also raised concerns about standardization and reliability, especially in high-stakes environments.

More recent research illustrates a move toward socially situated assessment practice, informed by sociocultural theory, systemic functional linguistics, and critical pedagogy (Papageorgiou & Bailey, 2019; Salaberry & Weideman, 2024). These frameworks situate assessment not as objective measurement but as a culturally and politically loaded act, demanding an ethical and interpretive stance—particularly in ESP contexts where learners' disciplinary and linguistic identities intersect.

From a metacognitive viewpoint, the initial assessment regimes offered little space for learner reflection, strategic planning, or adaptive thinking. It is necessary to consider this historical limitation to appreciate the degree to which the discipline has evolved and where there are still gaps in supporting metacognitive awareness in ESP–Engineering environments. Students were more likely to be passive receivers of test outcomes, which inhibited self-regulation and greater cognitive involvement with content.

The abstracted Table 1 below catalogues the features of this transformation identifying each milestone with their intrinsic peculiarities.

Table 01: A Historical Overview of Dominant Paradigms in Language Assessment

Time Period	Dominant Paradigm	Category	Sub-category	Details
1900s–1950s	Structuralist–Behaviorist	Theoretical Orientation	Language Philosophy	Language as discrete units; behaviorist learning
		Assessment Techniques	Methods	Multiple-choice, cloze, grammar tests
		Key Features	Focus	Discrete-point accuracy; objectivity
		Pedagogical Implications	Outcome	High reliability, limited communicative validity
1950s–1970s	Psychometric	Theoretical Orientation	Language Philosophy	Positivist stance; statistical focus
		Assessment Techniques	Methods	Standardized tests; statistical validation
		Key Features	Focus	Reliability, replicability, quantifiability
		Pedagogical Implications	Outcome	Objective results, limited pedagogical insight
1970s–1990s	Communicative	Theoretical Orientation	Language Philosophy	Communicative competence; functional view
		Assessment Techniques	Tasks	Oral interviews, role plays, integrated skills
		Key Features	Focus	Authenticity, context relevance

		Pedagogical Implications	Outcome	Learner agency, real-world performance
1990s–2010s	Sociocultural & Constructivist	Theoretical Orientation	Language Philosophy	Language as social practice; constructivist
		Assessment Techniques	Tasks	Performance-based, criterion-referenced tasks
		Key Features	Focus	Fairness, ethics, socially situated validity
		Pedagogical Implications	Outcome	Ethical framing, task realism
2010s–Present	Metacognitive (Critical Pedagogy & Post structural)	Theoretical Orientation	Language Philosophy	Critical, reflective, metacognitive focus
		Assessment Techniques	Methods	Dynamic assessment; AfL & AaL; digital feedback
		Key Features	Focus	Learner identity, reflection, agency
		Pedagogical Implications	Outcome	Contextual validity; metacognitive awareness

Note. Created by authors.

Integrative Nature of Language Assessment

Contemporary frameworks increasingly focus on the integrated nature of language ability, countering earlier tendencies to split listening, speaking, reading, and writing. The communicative language teaching (CLT) movement and task-based learning have led to tests that replicate everyday communication in which abilities co-occur naturally (Douglas & Abeywickrama, 2018).

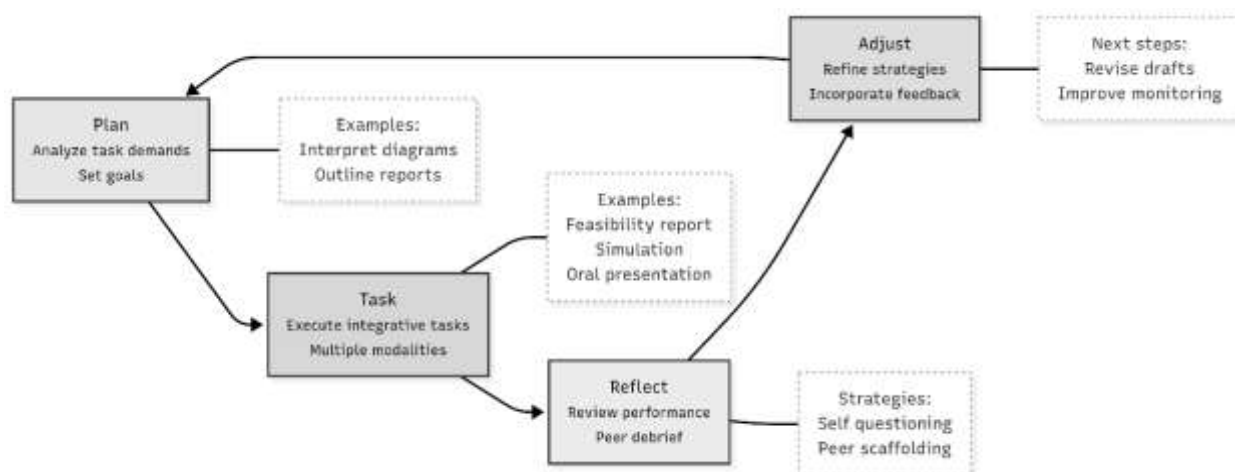


Figure 02: Integrative Assessment Cycle—Plan, Task, Reflect, Adjust.

Note. Created by authors

In ESP Engineering contexts, simulation, technical reports, and oral presentation performance-based measures compel students to induce multiple modalities into work contexts. For engineering education, that means activities such as diagramming interpretation, report writing on the feasibility, or technical briefs, such activities whose strong similarities have been found for workplace communication (Çelik, Çelik & Coombe, 2021).

In an integrated testing plot, not only do language skills interact but also the thinking and conceptualization of thoughts involving strategies to capitalize on achieving communicative outcomes. Huang & Jiang (2020) documented enhanced metacognitive effort among students engaged in integrative tasks. These empirical revelations as embedded in Fig.2 above attest to the efficacy of integrative tasks in stimulating planning, self-monitoring, and cross-modal reflection, those essential components of metacognition.

Moreover, integrative assessments provide learners with opportunities to develop cross-disciplinary problem-solving skills. By engaging with tasks that simulate real engineering scenarios—such as troubleshooting a circuit diagram or composing a report on fluid dynamics, students enhance their ability to transfer linguistic proficiency to authentic technical contexts.

Recent studies highlight that when learners participate in peer-led debriefing sessions after integrative tasks, they exhibit deeper conceptual understanding. These sessions foster critical reflection on both disciplinary content and language use, reinforcing metacognitive strategies such as self-questioning and peer scaffolding. This integration of skills not only enhances linguistic proficiency but also cognitive control wherein students must plan, monitor, and regulate their communicative strategy across modalities. These processes necessarily prompt metacognitive routines. By seeing, hearing, and reading simultaneously, and integrating all input in real time, learners become more flexible and more strategic in learning behavior, linking testing with the second aim of this review.

Dynamic and Alternative Assessment Paradigms

One of the key developments of recent years is the development of alternative assessments (AA) and dynamic assessments (DA). Contrasting with static, summative testing, these approaches prioritize learner agency, process awareness, and formative feedback. Portfolios, project work, reflection notebooks, and peer/self-assessment allow learners to demonstrate development over time and across a variety of contexts (Fulcher & Harding, 2022).

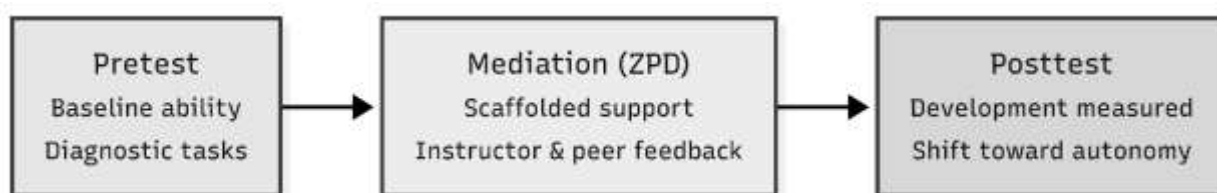


Figure 03: Dynamic Assessment in the Zone of Proximal Development—Pretest, Mediation, Posttest.

Note. Created by authors

Dynamic Assessment, rooted in Vygotsky's Zone of Proximal Development, emphasizes mediation in the assessment process, marking what the students can do with support but not independently (Poehner, 2008). This is very much aligned with Assessment as Learning (AaL) values, where the learner is both the object and the agent of the assessment. In Valsiner's (2001) own words, DA is a "present-to-future" journey (see Fig.3) that reveals embedded competence and directs individualized instruction.

DA has proven to be highly effective in L2 writing, speaking, and technical reporting contexts, where a scaffolding of feedback in real-time can be achieved for facilitating metacognitive development. In ESP–Engineering courses, think-aloud protocols, scaffolded revision sequences, and peer comment are techniques that serve as facilitators to triggering reflection and adaptive strategic thinking.

As Huang & Jiang (2020) highlight, the delivery of AA and DA requires rigorous teacher training, time investment, and school adaptability which are concerns that still remain particularly in resource-constrained South Asian contexts.

Table 02: Stages of Dynamic Assessment within the Zone of Proximal Development (ZPD).

Stage	Purpose	ESP–Engineering Examples	Metacognitive Effects
Pretest	Diagnose baseline ability	Initial draft of technical report, problem-solving task	Awareness of current performance level
Mediation	Scaffolded support (ZPD)	Think-aloud protocols, guided revision, peer comments	Reflection, adaptive strategy development
Posttest	Measure growth after mediation	Revised reports, oral presentations	Internalization of strategies, self-regulation

Note. Created by authors.

Recent implementations of dynamic assessment in engineering labs utilize immediate verbal feedback protocols, where instructors prompt students to articulate their reasoning while solving technical tasks. This interactive approach not only diagnoses gaps in conceptual understanding but also models metacognitive processes for learners.

Evaluation of interventionist DA protocols shows that graduated scaffolding that is moving from minimal to explicit prompts effectively builds students' self-regulation. As exemplified in Table 2. in detail, learners initially

rely on instructor cues but gradually internalize strategies for planning, monitoring, and self-correcting their written and oral technical communications.

Ethics, Ideology, and the Native Speaker Construct

Ethical issues have been further into the limelight in assessment studies. Beyond technical validity, teachers now consider issues of justice in assessment, especially in postcolonial and multilingual learning environments. Internal and external fairness are doubly guaranteed by Salaberry and Weideman (2024) such that their tests are not only procedurally valid but developmentally empowering.

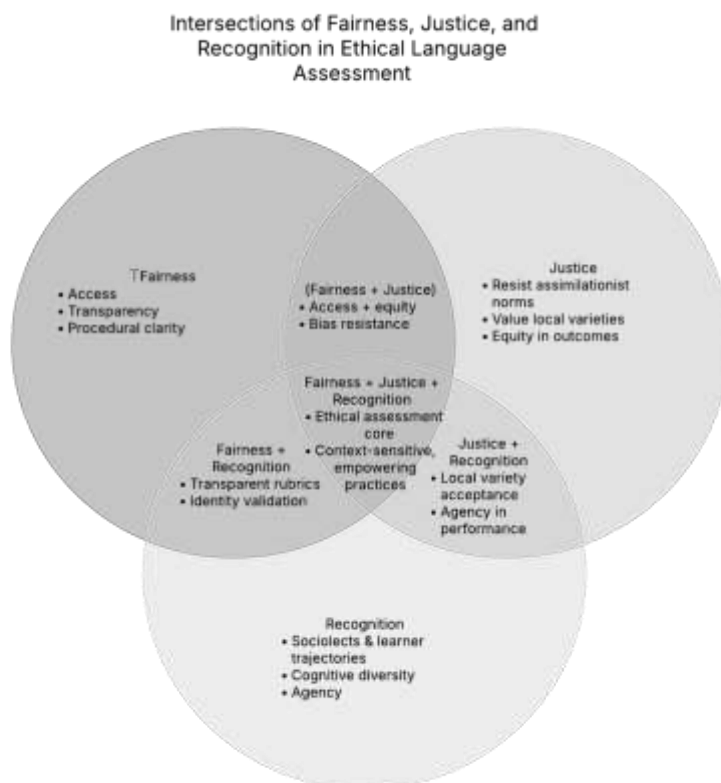


Figure 04: Dimensions of Ethical Language Assessment.

Note. Created by authors.

In ESP, native-speaker ideologies continue to distort perceptions of linguistic capacity. Pederson (2012) condemns the global preference for native-like English as a norm, a concept that excludes local varieties and learner identities. This is especially undesirable in engineering contexts where intelligibility, functionality, and clarity need to be prioritized over accent or idiomatic purity.

Implicitly sustaining native-speaker standards in examination frameworks carries the risk of further engendering language insecurity, deflating students' confidence and narrowing the syllabus. Reassessing rubrics, benchmarks, and marking schemes, there is consequently an imperative for realigning assessment towards functional sufficiency and local communication conventions—an ethical obligation in South Asian ESP courses.

Fig. 4 above interlaced the concerns of ethical language assessment exemplifying the multifaceted nature of this intricate web of facts. Fairness in assessment must be reconceptualized not only in terms of access and transparency but also in terms of recognition. Recognition includes validating learners' sociolects, cognitive processing strategies, and learning trajectories. It also implies resisting the assimilationist pressures that often accompany the global dominance of Anglophone norms. To counteract this, institutions must prioritize the development of local descriptors that reflect the communicative repertoire relevant to specific engineering contexts in multilingual societies.

Table 03: Core Ethical Dimensions Relevant to ESP Assessment Design, the Consequences of Neglecting them, and Context-Sensitive Strategies to Uphold Assessment Fairness and Integrity.

Ethical Dimension	Consequence of Neglect	Context-Sensitive Strategies
Fairness	Access inequality; lack of transparency	Ensure clear rubrics, open marking criteria, student consultation
Justice	Reinforcement of native-speaker bias; exclusion of local identities	Develop local descriptors; value intelligibility and functionality over accent
Recognition	Learner insecurity; reduced motivation	Validate sociolects, processing strategies, learner trajectories
Empowerment	Learners remain passive recipients; lack of ownership	Promote self-assessment, peer feedback, reflection notebooks
Contextual Relevance	Syllabi and rubrics disconnected from real-world ESP tasks	Engage stakeholders (industry, teachers, learners) in rubric design

As delineated in Table 3, this challenge invites broader stakeholder engagement from curriculum designers, policymakers, industry partners to learners themselves, in developing assessment criteria that are grounded in authentic, domain-specific communication rather than abstract native-like fluency. Doing so would allow learners to see their own linguistic identities reflected in what is being measured and ultimately foster a stronger sense of ownership and motivation. The principle of assessment justice therefore calls for critical interrogation of whose language practices are valued, whose are dismissed, and how tests may inadvertently reinforce structural inequalities.

While increasingly ideologized, this section is strongly metacognitive. When evaluation criteria acknowledge local language practices and learner identities, they empower learners to critically evaluate their language performance and trace it onto intended goals of their own. This feeling of ownership triggers self-regulation and reflective learning strategies that are central to metacognitive development.

Computational Tools and AI in ESP Examination

The growing integration of technology in language instruction has given birth to a new field of testing: digitally mediated, AI-guided, and learner-controlled. Within the context of ESP engineering workplaces, utilization of digital portfolios, learning management systems (LMS), and AI-driven feedback tools has provided entry points for learners to monitor, adapt, and trace their application of language in actual professional contexts. Tools such as Moodle, Google Classroom, and Turnitin support not only submission and marking but also embedded formative assessment to enable self-regulated learning.

Artificial intelligence tools such as Grammarly, Write & Improve, and feedback analytics in LMS dashboards support learners to track patterns of writing, gain instant formative feedback, and observe improvement over time. These affordances map well to metacognitive skills: planning, monitoring, and evaluating, and providing ESP students with agency over the self-management of language learning and disciplinary learning. In addition, mobile micro-assessment features can provide immediate feedback during lab report writing, technical presentation, and team meetings and facilitate situated learning within engineering contexts.

Even with their promise, such technologies must be used with pedagogical intent. They are not a replacement for teacher feedback but must be used to enrich an ecosystem of feedback. Digital assessment also carries ethical concerns, including data privacy, algorithmic bias, and differential access—of concern in the Global South. Institutional design, therefore, must be done to facilitate ethical, fair, and pedagogically grounded use of digital assessment tools.

Technology also supports metacognitive abilities by enabling visualization of progress, interpretation of feedback, and introspection about learning. Applications like Grammarly or Turnitin induce learners through cycles of iterative planning, self-monitoring, and revision, coordinating technological support with strategy training of the mind.

Assessment as a Catalyst for Metacognitive Development

ESP students, particularly those in engineering fields, must transcend surface linguistic accuracy to attain strategic competence: a skill to select, employ, and control language tools in the lived immediacy of problem-solving contexts. That necessitates an assessment framework that encourages self-regulation, scrutiny, and calibration—essential to metacognition principles (Flavell, 1979).

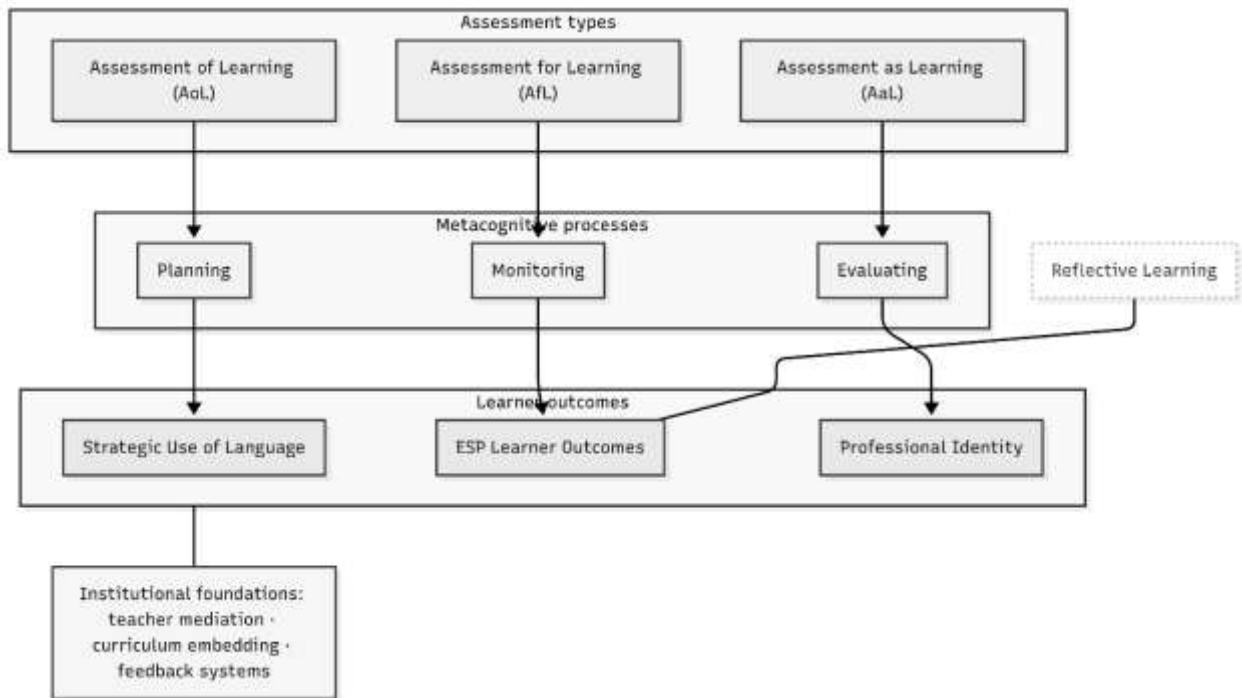


Figure 05: Assessment–Metacognition Alignment in ESP Contexts.

Research shows that Assessment as Learning (AaL) and Assessment for Learning (AFL) paradigms bear the most promise in shaping these skills (see Fig.5). Tools like reflective logs, self-assessment rubrics, and iterative project work facilitate learner agency, strategic planning, and recursive thinking.

Furthermore, flipped classroom designs and digital learning logs have been cost-effective and impactful strategies to scaffold ESP metacognition. Assignment logs, for instance, force learners to record plans of study, track progress, and reflect upon problems thereby granting immediate visibility of their problem-solving strategies and motivational routines.

To be effective, however, these strategies must be institutionally founded and embedded across the curriculum, not added on as add-on tools. Teachers must be trained to read learners' reflection, offer formative mediation, and modify instruction according to learners' cognitive profiles.

As it is clear, this aspect directly tackles the second purpose of this review. By locating assessment as a developmental, and not a judgmental, process, students have room to apply metacognitive processes that impact long-term academic and professional habits and such authentic examples of application can be traced in Table 4.

Table 04: A Comparative Analysis of Selected Assessment Types and Their Role in Fostering Metacognitive Strategies Among ESP–Engineering Learners, Demonstrating Cognitive Scaffolding Potential Across Modalities.

Assessment Type	Application in ESP	Metacognitive Strategy Fostered	Example in Engineering Context
Portfolios	Continuous collection of learner output	Reflection, self-monitoring	Design project documentation
Reflective Logs / Journals	Learners record plans, progress, problems	Self-questioning, recursive thinking	Weekly engineering lab logs
Self-Assessment Rubrics	Students evaluate their own performance	Strategic planning, calibration	Oral presentation rubric

Peer-Assessment	Peer review and critique of work	Social reflection, agency	Team project peer feedback
Iterative Project Work	Multiple draft–revision cycles	Regulation, adaptive strategies	Technical report drafts
Flipped Classroom Logs	Pre-/post-class reflection tasks	Monitoring, transfer of strategies	Digital assignment trackers

Washback and Impact: Shaping Learner Behaviors and Systems

Ultimately, the washback phenomenon also known as the impact of testing on learning and teaching, offers a dynamic perspective through which to consider the broader effect of test design. Fulcher & Harding (2022) maintain that washback is not an ancillary effect but a prominent feature of the effect of testing. Positive washback in ESP Engineering education happens when tests cause learners to engage reflectively with tasks, modify strategies, and absorb disciplinary discourse conventions.

Shrestha (2020) and Islam et al. (2021) illustrate the way grammar-filled, summative exams in South Asia produce adverse washback, discouraging learner autonomy and the adoption of rote learning approaches. In contrast, well-designed alternative assessments have been shown to produce developmental washback, eliciting strategic action, peer working, and the retention of languages in the long term.

Students who are exposed to cycles of iterative feedback—dialogic and actionable feedback—are likely to demonstrate higher motivation, self-efficacy, and discipline-specific fluency (Çelik, Çelik & Coombe, 2021). Hence, taking advantage of washback requires conscious design, constant calibration, and coordination with formative learning outcomes.

Washback effects give significant information regarding the function of assessment in shaping metacognitive engagement. Examinations that facilitate reflection, the integration of feedback, and adaptation of strategy constitute feedback cycle that enhances learners' self-perception and metacognitive control.

Reconceptualizing Validity in ESP Assessment

Later theory developments in language testing have reframed the validity construct as no longer narrowly statistical but a complex construct which is based on cognitive, social, and ethical constructs. Backing up this reconceptualization has been Messick's (1996) validity theory in its synoptic conception of bringing together content, criterion-related, and construct validity with consequential validity — with emphasis on the social consequence and ethics of assessment decisions.

For ESP for engineering, this shift demands a broader definition of what constitutes valid assessment. Not only do assessments need to test learners' knowledge but also how they use language in authentic, discipline-based contexts. Communicative validity therefore depends upon task realism, functional congruence with professional practice, and learners' ability to transfer strategic use of language into new situations. In addition, ecological validity, how much testing reflects real-world discourse communities, is particularly important in ESP engineering, where students must read technical reports, present, and work together to solve problems.

Tests must also be ethically valid, allowing for recognition of sociolinguistic variation in students, particularly in multilingual environments like Sri Lanka. The use of rigid rubrics or native-speaker norms can obscure students' progress and limit possibilities for development. Instead, evaluation should be a culturally responsive tool that adapts to students' linguistic backgrounds, cognitive profiles, and career goals. Consistently, ESP evaluation must evolve towards inclusive validity models that encompass fairness, impact, and pedagogical usefulness in addition to traditional psychometric stringency.

Adopting this expanded definition of validity not only enhances the validity of ESP tests but also supports learners' metacognitive engagement. If learners understand the rationale behind assessment design, criteria, and feedback, they are more likely to absorb reflective practices and become responsible for their learning activities and they remain features of assessment-literate and self-regulated learners in engineering.

Redefined validity to include pedagogical utility and student recognition enables metacognitive functions. As students view assessment as meaningful and fair, they will be more likely to engage reflectively, set goals, and revise their learning strategies accordingly tightening the metacognitive shift documented

Conclusions and Recommendations

This review has traced the historical and theoretical evolution of language testing and its intersection with metacognitive development in ESP–Engineering education. From early structuralist traditions to the emergence of dynamic, integrative, and ethically responsive paradigms, assessment has undergone significant transformation. Yet, in South Asian contexts, practices remain largely summative and certification-driven, often constraining learner agency and reflective growth.

The central conclusion is that metacognition must be embedded as a principle, not an adjunct, in ESP assessment design. By prioritizing reflection, planning, and self-assessment, assessment becomes a catalyst for strategic competence, disciplinary readiness, and professional autonomy. This shift not only supports lifelong learning but also aligns ESP testing with the evolving communicative demands of engineering education in multilingual environments.

The review identifies promising approaches—dynamic assessment, portfolios, reflective journaling, and integrative performance tasks—that are contextually meaningful and professionally relevant. However, their adoption requires institutional commitment to teacher assessment literacy and systemic embedding of alternative assessment practices, rather than their treatment as peripheral add-ons. Ethical considerations remain urgent, particularly the dismantling of native-speaker bias in rubrics and the development of locally grounded, intelligibility-focused descriptors.

Key recommendations include:

1. Embedding metacognitive scaffolding into assessment tasks through reflective logs, self-assessment rubrics, and feedback loops.
2. Diversifying formats to include dynamic, formative, and performance-based assessments that mirror authentic engineering communication.
3. Redesigning rubrics to value functional adequacy and intelligibility over native-speaker norms, consistent with global Englishes and local realities.
4. Prioritizing teacher professional development in assessment literacy, especially in designing metacognitively rich tasks.
5. Integrating alternative assessments into curricula at the design stage to ensure coherence and sustainability.

In doing so, this review contributes a regionally grounded, globally informed framework for re-engineering ESP assessment. It positions assessment not merely as a measurement tool but as a pedagogical catalyst for transformation—capable of shaping learner identity, fostering critical awareness, and promoting educational equity in the 21st-century ESP classroom.

References

- Çelik, S., Çelik, H., & Coombe, C. (Eds.). (2021). *Language assessment and test preparation in English as a foreign language (EFL) education*. Vizetek.
- Clavijo-Olarte, A., & Sharkey, J. (2019). Critical language teacher education in Colombia: Insights from teacher educators. In A. López-Gopar (Ed.), *International perspectives on critical pedagogies in ELT* (pp. 17–34). Palgrave Macmillan.
- Crookes, G. (2013). *Critical ELT in action: Foundations, promises, praxis*. Routledge.
- Davies, A. (2013). Fifty years of language assessment. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–6). Wiley-Blackwell. <https://doi.org/10.1002/9781118411360.wbcla127>
- Douglas, D., & Abeywickrama, P. (2018). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.
- Feuerstein, R., Rand, Y., & Rynders, J. E. (1988). *Don't accept me as I am: Helping "retarded" [sic] people to excel*. Plenum Press.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906–911.
- Freire, P. (1970). *Pedagogy of the oppressed*. Herder and Herder.
- Fulcher, G., & Harding, L. (Eds.). (2022). *The Routledge handbook of language testing* (2nd ed.). Routledge.
- Islam, M.S., Hasan, M.K., Sultana, S., Karim, A., Rahman, M.M. English language assessment

- in Bangladesh today: principles, practices, and problems. *Lang Test Asia* **11**, 1 (2021). <https://doi.org/10.1186/s40468-020-00116-z>
- Huang, R., & Jiang, L. (2020). Authentic assessment in Chinese secondary English classrooms: Teachers' perception and practice. *Educational Studies*, *46*(6), 1–14. <https://doi.org/10.1080/03055698.2020.1719387>
- Rahman, K.A., Seraj, P.M.I., Hasan, M.K. *et al.* Washback of assessment on English teaching-learning practice at secondary schools. *Language Testing in Asia*, *11*, 1–17. <https://doi.org/10.1186/s40468-021-00129-2>
- León Jiménez, M. A., López-Gopar, M. E., & Sayer, P. (2019). Critical language teacher education in Mexico: Transforming language teachers' beliefs and practices. In A. López-Gopar (Ed.), *International perspectives on critical pedagogies in ELT* (pp. 35–52). Palgrave Macmillan.
- Lidz, C. S., & Gindis, B. (2003). Dynamic assessment of the evolving cognitive functions in children. In A. Kozulin, B. Gindis, V. S. Ageyev, & S. M. Miller (Eds.), *Vygotsky's educational theory in cultural context* (pp. 99–116). Cambridge University Press.
- López-Gopar, M. E. (Ed.). (2019). *International perspectives on critical pedagogies in ELT*. Palgrave Macmillan.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*(3), 241–256. <https://doi.org/10.1177/026553229601300302>
- Norton, B., & Toohey, K. (2004). *Critical pedagogies and language learning*. Cambridge University Press.
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*, 422. <https://doi.org/10.3389/fpsyg.2017.00422>
- Papageorgiou, S., & Bailey, A. L. (Eds.). (2019). *Global perspectives on language assessment: Research, theory, and practice*. Routledge.
- Pederson, P. B. (2012). Ethics, competence, and professional issues in cross-cultural counseling. In P. B. Pederson, J. G. Draguns, W. J. Lonner, & J. E. Trimble (Eds.), *Counseling across cultures* (7th ed., pp. 3–28). SAGE Publications.
- Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Springer.
- Islam, M. S., Hasan, M. K., Sultana, S., & others. (2021). English language assessment in Bangladesh today: Principles, practices, and problems. *Language Testing in Asia*, *11*(1), 1. <https://doi.org/10.1186/s40468-020-00116-z>
- Rahman, M. S. (2017). The advantages and disadvantages of using qualitative and quantitative approaches and methods in language testing and assessment research: A literature review. *Journal of Education and Learning*, *6*(1), 102–112. <https://doi.org/10.5539/jel.v6n1p102>
- Rathnayake, W. M. P. Y. B. (2024). From tests to teaching: Exploring the washback effect in language pedagogy. *Journal of Desk Research Review and Analysis*, *2*(2), 145–158. <https://doi.org/10.4038/jdrra.v2i2.47>
- Salaberry, M. R., & Weideman, A. (Eds.). (2024). *Ethics and context in second language testing: Rethinking validity in theory and practice*. Routledge.
- Shrestha, P. N. (2020). *Dynamic assessment of students' academic writing: Vygotskian and systemic functional linguistic perspectives*. Springer.
- Valsiner, J. (2001). *The guided mind: A sociogenetic approach to personality*. Harvard University Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.