



From Benchmarking to Meaning-Making: A Conceptual Framework for Integrating SDG 4 into CEFR-Based Speaking Materials in Higher Education

Kalsum¹, Rahimah Saimin², Engku Muhammad Syafiq Engku Sharfuddin³

^{1,2,3} Universiti Sains Islam Malaysia, Malaysia

¹Institut Agama Islam Negeri Parepare, Indonesia

email: kalsum@iainpare.ac.id, 4241700@raudah.usim.edu.my

Abstract: *University speaking programmes increasingly adopt the Common European Framework of Reference for Languages (CEFR) to guide curriculum and assessment. However, existing implementation has largely emphasized assessment alignment, while the integration of sustainability and inclusion into CEFR-based speaking materials remains conceptually underdeveloped. This imbalance represents both a theoretical gap, as little attention has been paid to connecting CEFR's action-oriented philosophy with Education for Sustainable Development (SDG 4), and a practical gap, as materials developers lack systematic guidance for embedding these values into speaking instruction. Addressing this gap, this conceptual study develops an evidence-informed framework for integrating SDG 4 into CEFR-based speaking materials in higher education. The framework was constructed through a design-oriented conceptual analysis that systematically translated evidence synthesized from a previous systematic review and relevant theoretical literature into design requirements, which were subsequently consolidated into overarching design principles and organized into a coherent framework. The analysis generated five interrelated principles: thematic authenticity, descriptor-anchored task design, mediation as a vehicle for values, inclusive differentiation, and the separation of language assessment from content engagement. These principles were organized into a five-layer materials-development framework comprising descriptor literacy, SDG 4 thematic mapping, action-oriented task cycles, assessment integrity, and institutional support. Unlike the previously proposed Integrated CEFR Implementation Framework, which focuses on system-level implementation, the proposed framework contributes a product-level model to guide the development of sustainability-oriented CEFR-based speaking materials and is intended to serve as a conceptual foundation for future materials development and empirical validation in Indonesian higher education and comparable multilingual EFL contexts.*

Keywords: *CEFR, Higher Education, Materials Development, SDG 4, Speaking Instruction, Sustainable Education*

INTRODUCTION

The Common European Framework of Reference for Languages (CEFR) has become one of the most influential instruments in contemporary language education reform, and higher education institutions across Asia have increasingly

referenced it to structure their English programmes (Azman et al., 2021; Shak & Read, 2021). In speaking instruction particularly, CEFR offers an attractive promise: transparent proficiency levels, internationally comparable benchmarks, and a philosophy that treats learners as social agents who use language to act in the world (Council of Europe, 2001). Indonesian universities are part of this movement. English study programmes have begun expressing graduate competence in CEFR terms, commonly targeting the B1–B2 range for speaking, and aligning oral assessment procedures with can-do descriptors.

Speaking, however, remains a persistent challenge in Indonesian higher education. Recent studies published in this journal illustrate the difficulty from several angles. Suherman et al. (2026) report that the majority of non-English major students at an Indonesian private university perceive their proficiency as beginner level, with speaking-related anxiety and limited vocabulary among the dominant barriers. Sari et al. (2026) show that even when technology such as recorded WhatsApp video calls improves overall speaking performance, fluency remains the weakest aspect, constrained by fragmented practice opportunities. Asni et al. (2026), working in the multilingual classrooms of an Indonesian Islamic higher education institution, demonstrate that lecturers constantly negotiate between comprehension support and target-language exposure. Taken together, this local evidence suggests that speaking pedagogy in Indonesian universities requires not only proficiency benchmarks but carefully designed materials that scaffold learners toward sustained oral performance.

At the same time, higher education worldwide is expected to contribute to the 2030 Agenda for Sustainable Development, particularly SDG 4 on quality education. UNESCO (2017) calls for the integration of sustainability values, ethics, and global citizenship across disciplines rather than as isolated thematic additions, and inclusion has become a central criterion of educational quality (Ainscow, 2020). Language classrooms are, in principle, well positioned for this agenda: communicative tasks require content to talk about, and socially meaningful themes can supply that content while cultivating awareness. Yet in practice the two agendas rarely meet.

The empirical basis for this claim has recently been established. A PRISMA-guided systematic review of twelve empirical studies on CEFR-based speaking implementation in higher education (Kalsum et al., submitted) found that implementation is predominantly assessment-driven, that teachers' needs concentrate on descriptor and assessment literacy, and, most relevantly for the present study, that sustainability and inclusion remain peripheral rather than structurally embedded in current practice. The review answered the diagnostic question of what the field currently does. What it deliberately left open, and what no study has yet provided, is the prescriptive question that follows from its diagnosis: how, concretely, should speaking materials be designed so that CEFR alignment and SDG 4 values reinforce rather than ignore one another? Without an answer at the level of materials architecture, the review's call for sustainability-embedded pedagogy risks remaining an aspiration.

This conceptual study takes up that question. It pursues two research questions. First, what design principles for integrating SDG 4 content into CEFR-based speaking materials can be derived from the evidence synthesised in the prior review and from the wider literature on speaking pedagogy in higher education? Second, how can these principles be organised into a coherent materials-development framework, particularly for the Indonesian higher education context? The contribution is thus deliberately located downstream of the systematic review: where the review mapped the terrain and proposed a system-level implementation framework, the present study builds the product-level bridge between that evidence and the practical work of writing speaking materials. The remainder of the paper reviews the relevant literature, describes the conceptual method, presents the design principles and the proposed framework together with its differentiation from the implementation-level framework, and closes with implications and directions for empirical validation.

LITERATURE REVIEW

CEFR and the Action-Oriented Approach in Higher Education Speaking

CEFR was designed as a descriptive framework rather than a prescriptive syllabus. Its can-do descriptors articulate communicative performance across levels, and its action-oriented approach conceptualises learners as social agents who accomplish tasks under particular conditions and constraints (Council of Europe, 2001). Companion Volume extends this vision by elaborating the modes of reception, production, interaction, and mediation, positioning collaborative meaning-making as a core component of communicative competence (Council of Europe, 2018). In higher education speaking contexts, however, the framework has been taken up most visibly through assessment. Azman et al. (2021) document the alignment of a sustained monologue speaking test to CEFR levels in Malaysia, while Shak and Read (2021) align the language criteria of a group oral assessment in an English for occupational purposes classroom. Both studies illustrate the technical sophistication of descriptor-to-rubric alignment, and both belong to a wider pattern in which external frameworks are institutionalised through assessment standardisation before pedagogical innovation takes place (Fulcher, 2012; McNamara, 2001).

The assessment-first pattern is not inherently problematic; reliable scoring is a legitimate institutional need, and rater calibration protects learners from arbitrary judgement. The difficulty arises when alignment work absorbs most of the implementation effort, leaving task design, materials, and classroom interaction largely untouched. Curriculum research has long observed that policy frameworks tend to be translated into measurable outcomes before they are embedded in pedagogical practice (Priestley et al., 2015), and that the institutional structures surrounding teachers strongly condition whether reform reaches the classroom (Fullan, 2007). In language policy terms, proficiency frameworks in accountability-oriented systems often serve comparability and benchmarking functions first

(Ricento, 2015). For speaking materials development, this means that CEFR referencing can remain procedural unless deliberate design work reconnects descriptors with authentic communicative tasks.

SDG 4, Education for Sustainable Development, and Inclusion in Language Education

SDG 4 commits education systems to inclusive, equitable, quality education and lifelong learning for all, and Education for Sustainable Development (ESD) translates this commitment into learning objectives that span cognitive, socio-emotional, and behavioural domains (UNESCO, 2017). Two implications follow for language education. The first concerns content: ESD invites disciplines to treat sustainability themes as legitimate subject matter, which for communicative language teaching means that discussion, presentation, and problem-solving tasks can be built around issues such as educational access, skills for decent work, and global citizenship. The second concerns process: inclusion requires that instructional design accommodate learner diversity rather than presuppose homogeneity (Ainscow, 2020). Proficiency heterogeneity, unequal access to digital tools, and speaking anxiety, all extensively documented in Indonesian higher education (Sari et al., 2026; Suherman et al., 2026), are therefore not merely classroom management problems but equity concerns that materials design must anticipate.

Within CEFR scholarship, these agendas find natural anchor points. The action-oriented approach already frames language use as socially situated action, and mediation foregrounds the collaborative construction of meaning across linguistic and cultural boundaries (Council of Europe, 2018). Conceptually, then, a CEFR-based speaking programme could carry ESD content without distorting either framework. Empirically, however, the systematic review on which this study builds found the connection to be thin and unevenly conceptualised (Kalsum et al., submitted), which is precisely the design problem the present article addresses.

Task, Technology, and Innovation in University Speaking Instruction

A further strand of research documents the pedagogical resources available for speaking materials design. Task-based learning combined with graphic organiser techniques has been shown to develop student teachers' listening and speaking competency (Kositchaivat, 2025), and digital portfolios have enhanced EFL speaking performance and engagement in higher education (Cabrera-Solano, 2020). Blended and bichronous configurations support short oral presentations among undergraduates (Manipatruni et al., 2024), while video podcast creation develops both speaking and digital competence in English for specific purposes settings (Brega & Kruglyakova, 2025). More recent work explores artificial intelligence, from conversational teacher-bots that expand speaking practice

opportunities (Benaissa et al., 2025) to AI-driven assessment whose effects on motivation, anxiety, and proficiency are compared with traditional examiner-led formats (Mirsanjari, 2025). Creative multimodal tasks such as audio description have also been proposed to widen the communicative repertoire (Ogea Pozo & Ruiz Espejel, 2024).

This strand matters for the present study in two ways. It demonstrates that the field possesses a rich inventory of task types and delivery modes through which CEFR descriptors could be operationalised communicatively; and it repeatedly surfaces equity considerations, such as connectivity constraints and anxiety, that any sustainability-oriented framework must address rather than assume away. In the Indonesian context, Sari et al. (2026) provide a telling illustration: mobile-mediated speaking practice improved performance yet was undermined by unstable internet access, an infrastructural inequality that the authors explicitly link to policy responsibility. Materials development that claims alignment with SDG 4 cannot treat such conditions as background noise.

RESEARCH METHOD

This study employed a design-oriented conceptual method: it develops design principles and a framework from an existing evidence base rather than generating new empirical data. Conceptual and framework-building studies of this kind are recognized contributions in review methodology when they move systematically from synthesized evidence to actionable constructs and make each derivation step transparent (Snyder, 2019; Xiao & Watson, 2019). Three sources of evidence and theory were used. The first and primary source is the completed systematic review of twelve empirical studies on CEFR-based speaking implementation in higher education (Kalsum et al., submitted), which followed PRISMA 2020 procedures (Page et al., 2021) across Scopus, Web of Science, and ERIC, and whose findings on teachers' needs, operationalization models, sustainability orientations, and structural gaps constitute the diagnostic ground of the present work. The second source comprises the canonical CEFR and ESD documents that define the constructs being connected (Council of Europe, 2001, 2018; UNESCO, 2017). The third source is recent empirical work on speaking instruction in Indonesian higher education (Asni et al., 2026; Sari et al., 2026; Suherman et al., 2026), used to test the contextual plausibility of each design decision.

The analysis proceeded in three steps. In the first step, principle derivation, each gap pattern reported in the prior review was reformulated as a design requirement by asking what a materials developer would need to do differently for the gap not to be reproduced. For example, the review's finding that sustainability appears mainly as an implicit equity concern was reformulated as the requirement that inclusion be designed into task scaffolding rather than handled reactively in the classroom. Requirements that recurred across gap patterns were consolidated into candidate principles, which were then checked against the wider speaking-pedagogy literature for theoretical support and against the Indonesian studies for contextual fit.

In the second step, framework construction, the consolidated principles were organised into layers through iterative mapping, ordered by logical dependency: a layer was placed below another when decisions at the lower layer constrain decisions at the upper one. Each layer was checked back against the review's evidence categories to ensure grounding. In the third step, differentiation analysis, the resulting construction was compared systematically with the system-level Integrated CEFR Implementation Framework proposed in the prior review, to establish the distinct scope, audience, and function of each and to specify how the two relate. The framework presented here should therefore be read as an evidence-informed conceptual construction whose empirical validation, through the development and trialling of actual materials, is identified as the necessary next step and is being pursued in a separate design and development study (Kalsum et al., 2026).

FINDINGS AND DISCUSSION

The Design Space: What Review Evidence Establishes

Before principles can be derived, the design space must be delimited by what the evidence shows. The prior review identified three distinguishable stances toward sustainability-related content across its corpus of twelve studies (Kalsum et al., submitted); Table 1 summarises this distribution, which the present study takes as its point of departure rather than as a new finding.

Table 1. *Stances toward sustainability content in CEFR-based speaking research (summarised from Kalsum et al., submitted)*

Stance	Studies	Characteristic practices	Representative sources
Explicit integration	3 of 12	Speaking tasks designed around socially relevant themes (global citizenship, ethical communication, creativity and access); differentiated task pathways consistent with inclusive design	Ogea Pozo & Ruiz Espejel (2024)
Implicit equity orientation	5 of 12	Inclusion treated as practical concern: fairness in assessment, anxiety reduction, digital access, support for lower-proficiency learners;	Mirsanjari (2025); Brega & Kruglyakova (2025); Manipatruni et al. (2024); Benaissa et al. (2025)

Stance	Studies	Characteristic practices	Representative sources
		sustainability not named as a construct	
Absence	4 of 12	Focus confined to descriptor-to-rubric alignment, scoring calibration, and validation; content themes incidental to the assessment procedure	Azman et al. (2021); Shak & Read (2021)

What the present study adds to this distribution is a design-oriented re-reading. Viewed diagnostically, as the review viewed it, the distribution shows that sustainability is peripheral. Viewed through a design lens, however, the three stances appear not as fixed positions but as points on a developmental path, and the distance between them is shorter than the scarcity of explicit studies implies. Studies in the implicit group already possess the raw material for explicit integration: they manage heterogeneity, worry about fairness, and scaffold anxious learners. Naming these practices as inclusion, and selecting communicative themes deliberately, would move such programmes into the explicit group without altering their instructional machinery. Even the absence group is closer than it looks; its rigorous descriptor work is exactly the foundation that thematically rich materials require. The design problem is therefore not to invent sustainability-embedded speaking pedagogy from nothing, but to supply the connecting architecture that lets existing practices travel this path. The five principles that follow constitute that architecture.

Five Design Principles for Sustainability-Embedded, CEFR-Based Speaking Materials

First, thematic authenticity: sustainability themes should function as communicative content, not tested content. The explicit-integration studies succeed precisely because socially relevant themes supply something worth talking about, while language remains the object of instruction and assessment (Ogea Pozo & Ruiz Espejel, 2024). The moment sustainability knowledge becomes an assessment target, the speaking programme drifts toward content examination and the affective benefits of meaningful topics are lost. Materials should therefore frame SDG 4 issues, such as access to education or skills for the future, as discussion-worthy problems, never as declarative knowledge to be recited.

Second, descriptor-anchored task design: every task should be traceable to specific CEFR descriptors at the targeted level. The alignment studies demonstrate the discipline this requires, distinguishing adjacent levels such as B1 and B2 and holding tasks to level-appropriate demands (Azman et al., 2021; Shak & Read, 2021). Sustainability-embedded materials inherit this discipline; thematic richness is no excuse for level drift. In practice this means specifying, for each task, the production and interaction descriptors it operationalises, and grading input texts so that content complexity does not outrun linguistic accessibility.

Third, mediation as a vehicle for values: the collaborative modes elaborated in the Companion Volume (Council of Europe, 2018) are the natural carriers of sustainability content. Tasks in which learners relay, summarise, negotiate, and build shared understanding, such as group problem-solving around an educational access scenario, simultaneously exercise mediation competences and enact the collaborative dispositions ESD seeks to cultivate (UNESCO, 2017; Kositchaivat, 2025). The prior review found mediation to be the least operationalised mode (Kalsum et al., submitted); embedding it through value-laden collaborative tasks addresses two gaps with one design decision.

Fourth, inclusive differentiation: materials should anticipate heterogeneity rather than presuppose uniform proficiency. The implicit-equity studies document the consequences of ignoring this: anxiety suppresses participation, connectivity constraints fragment practice, and lower-proficiency learners disengage (Mirsanjari, 2025; Sari et al., 2026; Suherman et al., 2026). Concretely, tasks require tiered scaffolding, such as layered sentence frames, planning time, and low-stakes rehearsal cycles, so that the same communicative task remains accessible across the B1–B2 span, consistent with inclusive education principles (Ainscow, 2020).

Fifth, assessment separation: language performance and content engagement must be assessed, or observed, through separate instruments. CEFR analytic criteria evaluate range, accuracy, fluency, coherence, and interaction; engagement with sustainability themes is better captured through reflective instruments such as journals or structured self-report, which respect the non-linguistic nature of the construct. Conflating the two would compromise both the validity of the language assessment (McNamara, 2001) and the sincerity of learner reflection. The corpus reviewed previously offers a cautionary contrast: AI-driven and examiner-led assessment formats already struggle to balance reliability with learner affect (Mirsanjari, 2025); loading value judgements onto the same instruments would aggravate the strain.

A Layered Materials-Development Framework

The second research question asked how these principles can be organised into a coherent framework. Figure 1 presents the proposed architecture, which arranges materials development as five interdependent layers.

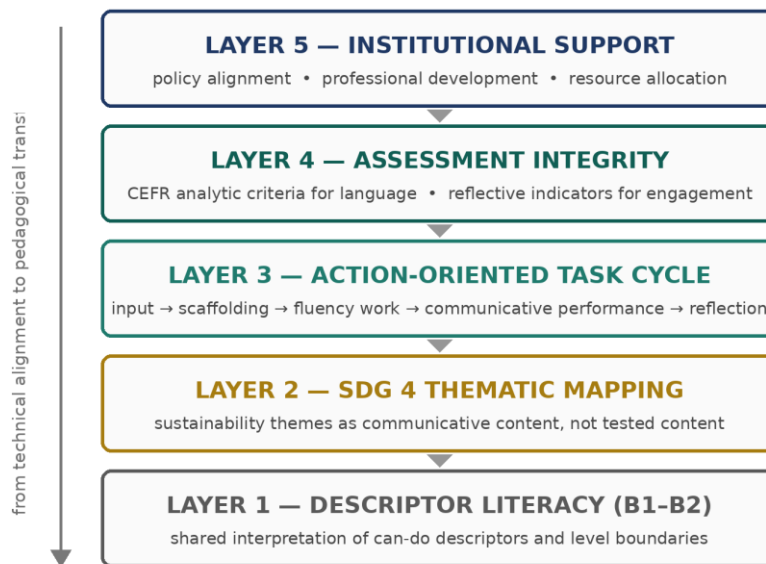


Figure 1. A layered materials-development framework for sustainability-embedded, CEFR-based speaking instruction

The foundation layer, descriptor literacy, ensures that developers and teachers share an interpretation of the targeted can-do descriptors and level boundaries; without it, every subsequent decision is unanchored, a risk the alignment literature documents extensively (Azman et al., 2021; Shak & Read, 2021; Fulcher, 2012). The second layer maps SDG 4 themes onto the programme as communicative content, selecting issues that are discussion-worthy, locally resonant, and level-appropriate. The third layer converts descriptors and themes into action-oriented task cycles, sequences that move from thematic input through language scaffolding and fluency work to communicative performance and reflection, drawing on the task and technology repertoire the field already possesses (Kositichaiwat, 2025; Cabrera-Solano, 2020; Manipatruni et al., 2024). The fourth layer protects assessment integrity by running CEFR analytic criteria for language alongside reflective indicators for engagement, in line with the separation principle. The outermost layer, institutional support, acknowledges that materials survive only within supportive structures of policy, professional development, and resources (Fullan, 2007; Priestley et al., 2015; Ricento, 2015); the vertical arrow in the figure signals that the whole construction represents a movement from technical alignment toward pedagogical transformation.

A clarification of scope is necessary here, because the prior review proposed a framework of its own, the Integrated CEFR Implementation Framework (Kalsum et al., submitted), and the two constructions must not be conflated. The implementation framework operates at the system level: its unit of analysis is the institution, its audience comprises policy-makers, programme leaders, and quality assurance bodies, and its function is to describe the conditions under which CEFR adoption becomes systemic rather than procedural. The framework proposed here operates at the product level: its unit of analysis is the speaking material itself, its audience comprises materials writers, module developers, and lecturers, and its function is prescriptive, specifying the design commitments a sustainability-embedded speaking module must honour. The relationship between the two is nested rather than parallel. The present framework operationalises, at materials scale, what the implementation framework's pedagogical layers call for at institutional scale; two of its layers have no counterpart in the implementation framework, namely the explicit SDG 4 thematic mapping layer and the dual-track treatment of assessment, and these constitute its distinct contribution. An institution could adopt the implementation framework without knowing how to write a single unit of material; this framework supplies exactly that missing know-how.

Two further features distinguish the framework from a generic materials-development checklist. The first is directionality: lower layers constrain upper ones, so a programme cannot compensate for weak descriptor literacy with elaborate themes, nor for missing institutional support with well-crafted tasks. The second is the explicit dual-track treatment of assessment, which resolves the tension, visible throughout the reviewed corpus, between the reliability imperative of institutions and the formative, affect-sensitive needs of learners (Mirsanjari, 2025; Sari et al., 2026).

For Indonesian higher education, and Islamic higher education institutions in particular, the framework has immediate applicability. The local studies reviewed earlier describe exactly the conditions the framework is designed for: multilingual classrooms in which lecturers balance comprehension support with target-language exposure (Asni et al., 2026), learner populations concentrated at lower proficiency levels with pronounced speaking anxiety (Suherman et al., 2026), and technology-mediated practice constrained by infrastructure (Sari et al., 2026). A B1–B2 speaking programme built on this framework would select SDG 4 themes with strong local resonance, such as educational access in remote regions, scaffold them through tiered frames and fluency cycles, assess language through CEFR analytic rubrics, and capture learners' developing awareness through reflective journals, while the institution commits to lecturer development in descriptor and assessment literacy. A prototype module built on precisely these commitments is being developed and reported separately (Kalsum et al., 2026). In this

configuration, CEFR referencing stops being an exercise in benchmarking and becomes an architecture for meaning making.

CONCLUSION

This study set out to build the bridge that a recent systematic review showed to be missing between CEFR-referenced speaking instruction and the educational values of SDG 4. Taking the review's diagnosis as its point of departure, that implementation is assessment-driven and sustainability peripheral (Kalsum et al., submitted), the study derived five design principles, thematic authenticity, descriptor-anchored task design, mediation as a vehicle for values, inclusive differentiation, and assessment separation, and organised them into a five-layer materials-development framework spanning descriptor literacy, SDG 4 thematic mapping, action-oriented task cycles, assessment integrity, and institutional support.

The contribution is deliberately practical and deliberately scoped. The framework gives materials developers, curriculum designers, and lecturers a sequenced set of design commitments through which speaking programmes can serve proficiency development and quality-education values simultaneously, without conflating the two in assessment. It complements, rather than repeats, the system-level implementation framework of the prior review: where that framework tells institutions what conditions to build, this one tells developers what materials to build within those conditions. Together they trace a route from benchmarking to meaning-making.

The limitations are those of the design. The framework is an evidence-informed conceptual construction, grounded in a small corpus and interpretive derivation, and it has not yet been empirically tested. Future research should therefore validate it through design and development studies that produce and trial actual materials, examine learner outcomes across the B1–B2 span, and investigate longitudinally whether sustainability-embedded speaking programmes sustain both proficiency gains and learner engagement. Work of this kind is currently underway in the Indonesian Islamic higher education context through an ADDIE-based prototype study (Kalsum et al., 2026), for which the framework presented here serves as the conceptual foundation, and it is offered equally as a reference for other institutions pursuing the same integration.

ACKNOWLEDGEMENTS

This article is part of the first author's doctoral research at the Faculty of Major Language Studies, Universiti Sains Islam Malaysia (USIM), conducted in collaboration with the State Islamic Institute of Parepare (IAIN Parepare), Indonesia. The systematic review on which this study builds is reported separately and is currently under editorial consideration; the relationship between the two manuscripts has been disclosed to the editors of both.

REFERENCES

- Ainscow, M. (2020). Promoting inclusion and equity in education: Lessons from international experiences. *Nordic Journal of Studies in Educational Policy*, 6(1), 7–16. <https://doi.org/10.1080/20020317.2020.1729587>
- Asni, Y., Hasmi, & Amalia, R. N. (2026). Between flexibility and language exposure: Translanguaging and language immersion perceptions in university English classrooms. *Indonesian TESOL Journal*, 8(1), 36–47.
- Azman, H., Othman, Z., Mohd. Shamsuddin, C., Wahi, W., Abd Aziz, M. S., Wan Mohamad, W. N., Othman, S., & Mohd Amin, M. H. (2021). Relating a sustained monologue speaking production test to CEFR: Towards alignment. *Pertanika Journal of Social Sciences & Humanities*, 29(S3), 385–400. <https://doi.org/10.47836/pjssh.29.S3.20>
- Benaissa, M., Attou, Y., Seddik, M., & Nfissi, A. (2025). Can AI-based teacher-bots (T-bots) enhance students' speaking skills? A case study of “My AI” on Snapchat in Moroccan higher education. *Arab World English Journal, Special Issue on CALL*, 11, 144–159. <https://doi.org/10.24093/awej/call11.9>
- Brega, O. N., & Kruglyakova, G. V. (2025). Digital competence development while video podcast creation in ESP. *Perspectives of Science and Education*, (2), 210–227. <https://doi.org/10.32744/pse.2025.2.14>
- Cabrera-Solano, P. (2020). The use of digital portfolios to enhance English as a foreign language speaking skills in higher education. *International Journal of Emerging Technologies in Learning (iJET)*, 15(24), 159–175. <https://doi.org/10.3991/ijet.v15i24.15103>
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume with new descriptors*. Council of Europe Publishing.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- Fullan, M. (2007). *The new meaning of educational change* (4th ed.). Teachers College Press.
- Kalsum, K., Saimin, R., & Engku Sharfuddin, E. M. S. (submitted). Teachers' needs in developing CEFR-based speaking materials: A systematic literature review toward sustainable and inclusive higher education [Manuscript submitted for publication]. Faculty of Major Language Studies, Universiti Sains Islam Malaysia.
- Kalsum, K., Saimin, R., & Engku Sharfuddin, E. M. S. (2026). Designing CEFR-aligned content-based speaking materials using SDG 4 quality-education themes in Islamic higher education: An ADDIE-based prototype as an innovative TESL practice [Paper presentation]. *International Conference on Teaching English and Language Learning (ICTELL 2026)*.
- Kositchaivat, S. (2025). The effects of task-based learning and graphic organizer

- techniques on developing student teacher professional competency in listening and speaking skills. *LEARN Journal: Language Education and Acquisition Research Network*, 18(1), 859–882. <https://doi.org/10.70730/NESC6564>
- Manipatruni, V. R., Siva Kumar, N., Karim, M. R., & Banu, S. (2024). Synthesis of blended learning and bichronous learning in improving undergraduates' English-speaking skills through short presentations. *XLinguae*, 17(1), 107–122. <https://doi.org/10.18355/XL.2024.17.01.08>
- McNamara, T. (2001). Language assessment as social practice. *Language Testing*, 18(4), 333–349. <https://doi.org/10.1177/026553220101800402>
- Mirsanjari, Z. (2025). AI-driven vs. traditional language assessment: Effects on Iranian EFL learners' motivation, anxiety, and proficiency in a high-stakes exam context. *Language Testing in Asia*, 15, 70. <https://doi.org/10.1186/s40468-025-00395-4>
- Ogea Pozo, M. del Mar, & Ruiz Espejel, B. (2024). Enhancing foreign-language creativity: A didactic approach through audio description. *Complutense Journal of English Studies*, 32, e97137. <https://doi.org/10.5209/cjes.97137>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Priestley, M., Biesta, G., & Robinson, S. (2015). *Teacher agency: An ecological approach*. Bloomsbury.
- Ricento, T. (2015). Language policy and planning in higher education. In T. Ricento (Ed.), *Language policy and political economy*. Oxford University Press.
- Sari, D. F., Wati, N., & Silviyanti, T. M. (2026). The analysis of the EFL students' speaking performances in learning English by recorded WhatsApp video call. *Indonesian TESOL Journal*, 8(1), 20–35.
- Shak, P., & Read, J. (2021). Aligning the language criteria of a group oral test to the CEFR: The case of a formal meeting assessment in an English for occupational purposes classroom. *Pertanika Journal of Social Sciences & Humanities*, 29(S3), 133–156. <https://doi.org/10.47836/pjssh.29.S3.08>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Suherman, L. O. A., Akfan, S. S., Hamka, N., Hasan, A., Riniati, W. O., & Arba'a, N. A. (2026). English learning profiles of non-English major students: A comparative study between Accounting and Management programs. *Indonesian TESOL Journal*, 8(1), 1–19.
- UNESCO. (2017). *Education for sustainable development goals: Learning objectives*. UNESCO Publishing.
- Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, 39(1), 93–112. <https://doi.org/10.1177/0739456X17723971>