Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Debriefing and Structured Reflection after Communicative Tasks in Second Language and ESP Learning: A PRISMA-Guided Systematic Review with VOSviewer Bibliometric Mapping (2000–2026)

Bahodir Ismoilovich Shokirov

openai@ajou.uz

AJOU University in Tashkent, Uzbekistan

Charos Akbarovna Sharipova

Uzbekistan State World Languages University, Uzbekistan

Bokhodir Makhkamovich Kuchkarov

AJOU University in Tashkent, Uzbekistan

Research Article

Keywords: debriefing, reflection, post-task, self-assessment, task-based language teaching, simulation, English for specific purposes, systematic review, bibliometric analysis, VOSviewer

Posted Date: October 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11518342/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Background.

Simulations, role plays, and other communicative tasks are central to second language (L2) and English for Specific Purposes (ESP) teaching. Experiential-learning and simulation theory treat the reflective phase that follows such tasks (debriefing, guided reflection, self-review) as the place where experience becomes learning. No review has yet synthesised what is known about this post-performance phase in L2 learning.

Methods.

Following PRISMA 2020, three Boolean queries were run in OpenAlex (2000–2026). A documented rule-based screen produced a bibliometric corpus of 888 records, which was mapped in VOSviewer 1.6.21 through term co-occurrence, overlay, and co-citation analyses. From this corpus, 405 candidates were screened against eligibility criteria for empirical studies in which L2 learners engaged in a structured reflective phase after an oral or interactive task. Included studies were coded for context, design, reflective modality, and outcomes, appraised with an MMAT-informed rubric, and synthesised narratively by modality.

Results.

Forty-three studies (46 reports) met the criteria. They used five reflective modalities: structured debriefing (10 studies), post-task transcription (10), guided self-reflection or self-assessment (15), recording-based self-review (9), and peer or multi-source feedback with reflection (13). The most consistent controlled evidence concerned accuracy gains from transcription and guided reflection between task repetitions, although one randomised study found no added benefit. Evidence for structured debriefing of simulations and role plays rested mainly on qualitative, survey, and action-research designs. Only 12 of the 88 corpus records that mentioned debriefing described it as a pedagogical phase after an L2 task. Co-citation analysis showed that the field draws on experiential, task-performance, task-based-pedagogy, and interaction/noticing traditions, but not on the simulation-debriefing literature.

Conclusions.

Post-performance reflection in L2 learning is promising, but the evidence is fragmented and methodologically uneven. A five-phase reflective cycle (performance, evidence capture, structured reflection, calibration, and feedforward) is proposed to guide practice and a research agenda that prioritises controlled comparisons of debriefing formats, interactional and professional outcomes, and transparent reporting.

1. INTRODUCTION

Communicative tasks such as role plays, simulations, presentations, and negotiations have become a mainstay of second language (L2) and English for Specific Purposes (ESP) instruction, because they ask learners to use language to get something done (Ellis 2003; Willis 1996). Simulation has a long history in language teaching (Jones 1982), and a recent systematic review of role play in EFL and ESL settings stresses that its effects depend on deliberate design, preparation, and reflective debriefing (Kaygısız and Akar 2025). A recent analysis of ESP

classrooms makes the same point, arguing that debriefing is undervalued in ESP even though scenario-based activities are increasingly common (Šulovská 2026).

The rationale for a post-performance reflective phase comes from several traditions. In experiential-learning theory, concrete experience becomes learning only when it is followed by reflective observation and abstract conceptualisation (Kolb 1984; Knutson 2003). In simulation and gaming, debriefing is treated as the phase in which participants make sense of what happened and transfer it to future action (Lederman 1992; Crookall 2010). Health-professions education has formalised debriefing into structured, evidence-based conversations and has shown that how debriefing is conducted matters as much as whether it occurs (Fanning and Gaba 2007; Rudolph et al. 2006; Eppich and Cheng 2015; Sawyer et al. 2016). In second language acquisition (SLA), the cognitive case rests on noticing: learners must attend to the gap between their output and target forms for that output to drive acquisition (Schmidt 1990; Long 1996). Task-based research shows that the post-task phase, and the anticipation of a post-task activity, can redirect learners' attention towards accuracy (Skehan 1998; Foster and Skehan 2013). Research on feedback and self-regulation adds that learners benefit when feedback is processed, compared with internal standards, and converted into plans for the next performance (Hattie and Timperley 2007; Nicol and Macfarlane-Dick 2006; Zimmerman 2002; Panadero, Jonsson, and Botella 2017).

Despite these converging arguments, the empirical literature on what L2 learners do after a communicative performance is scattered. It appears under labels as different as debriefing, post-task transcription, self-reflection, self-assessment, video self-review, and peer feedback, and in journals ranging from SLA and CALL to medical education. No synthesis has integrated these strands, and it is not known what reflective formats have been tried, what the designs are, what the outcomes are, or how consistently. The gap is not only in theory, but also in practice. Few teachers who develop course simulations or role plays have any research-supported advice on how to design the reflective part of the course that comes next.

This study fills this gap by conducting a PRISMA-guided systematic review (Page et al. 2021) and a bibliometric mapping using the VOSviewer tool of the literature surrounding this. (van Eck and Waltman 2010) The bibliometric component identified the review in the thematic and intellectual space of the field and assesses the extent to which the scholarship of simulation-debriefing has penetrated L2 research. The empirical evidence is then synthesised in the systematic review. The following four research questions help guide the study:

RQ1. What are the thematic structure, intellectual base, and evolution of research on reflection and debriefing in L2 communicative learning (2000–2026)?

RQ2. Which forms of post-performance reflection have been studied empirically, with which learners, in which contexts, and with what designs?

RQ3. What outcomes are associated with each reflective modality, and how consistent and methodologically robust is the evidence?

RQ4. What design principles and research priorities follow for L2 and ESP pedagogy?

2. METHOD

2.1 Design and reporting

The study combined a bibliometric mapping (RQ1) with a systematic review and narrative synthesis (RQ2–RQ4). Reporting follows PRISMA 2020 (Page et al. 2021), and the bibliometric procedures follow established guidance (Donthu et al. 2021; Zupic and Čater 2015). The review protocol was not prospectively registered; all queries, screening rules, decisions, and coding are provided as supplementary files so that the review can be fully reproduced.

2.2 Information source

OpenAlex (Priem, Piwowar, and Orr 2022) was used as the single bibliographic source. Web of Science and Scopus were not accessible for bulk export during the review. OpenAlex is an open index whose reference coverage is broadly comparable to the commercial databases for jointly indexed publications (Culbert et al. 2025). It is also more inclusive of the regional and open-access journals in which much classroom L2 research appears, and VOSviewer supports it directly. The search strings are given in Table 1 so that the review can be replicated in Web of Science and Scopus.

2.3 Search strategy

Three complementary queries were run on titles and abstracts on 27 October 2026, limited to 2000–2026, English, and the document types article, review, and book chapter (Table 1). Q1 targeted debriefing in L2 contexts. Q2 targeted post-task and guided reflection after oral or interactive tasks. Q3 targeted reflection, debriefing, or metacognition in simulation, role play, and experiential L2 learning. The queries returned 177, 475, and 531 records; after 61 cross-query duplicates were removed, 1,122 unique records remained.

Table 1
Information source, search strings, and eligibility criteria

Element	Specification
Source and date	OpenAlex REST API, title_and_abstract.search; 27 October 2026; 2000–2026; article, review, book chapter; English
Q1 (n = 177)	(debrief OR debriefing OR debriefed) AND (EFL OR ESL OR L2 OR "second language" OR "foreign language" OR "English for specific purposes" OR "language learning" OR "language learners" OR "language teaching" OR "language education")
Q2 (n = 475)	("post-task reflection" OR "guided reflection" OR "structured reflection" OR "reflective discussion" OR "self-reflection" OR "reflective feedback" OR "post-task") AND (speaking OR oral OR simulation OR "role play" OR "task-based" OR interaction) AND (EFL OR ESL OR L2 OR "second language" OR "foreign language" OR "English for specific purposes")
Q3 (n = 531)	(simulation OR simulations OR "simulation-based" OR "role play" OR "role-play" OR "experiential learning") AND (reflection OR reflective OR debriefing OR "reflective practice" OR metacognitive) AND (EFL OR ESL OR "second language" OR "foreign language" OR "English for specific purposes" OR "language learning")
Population	Inclusion: L2/foreign-language learners of any age, including ESP and professional-purpose learners. Exclusion: teachers or teacher candidates as reflecting agents
Phenomenon	Inclusion: designed post-performance reflective phase (debriefing, guided/structured reflection, post-task transcription, self-assessment/self-review, reflectively processed peer or multi-source feedback) after an oral, interactive, or communicative task. Exclusion: reflection used only as a research instrument; exclusively written tasks
Outcomes	Learner outcomes, processes, or perceptions related to the reflective phase
Study type	Inclusion: empirical peer-reviewed articles and scholarly chapters. Exclusion: conceptual papers, reviews, conference proceedings
Note. Web of Science/Scopus equivalents (TS= / TITLE-ABS-KEY) of the three strings are provided in Supplementary File S1 for replication.	

2.4 Bibliometric corpus and analyses

The 1,122 records were screened with three concept blocks applied to title and abstract: a reflection term (e.g., debrief, *reflect*, post-task, metacogniti, *self-assess*), a task or communication term, and an L2 term. Records without an abstract of at least 100 characters, duplicated titles, and records whose OpenAlex primary topic fell outside the language and education domains were excluded (234 in total). The resulting corpus of 888 records (16,014 cited references) was analysed in VOSviewer 1.6.21, run from the command line so that every parameter is recorded. Term co-occurrence analysis used title–abstract text, binary counting, a 363-entry cleaning thesaurus, and a minimum of 25 occurrences, which yielded 101 terms. Clustering used resolution 1.00 and a minimum cluster size of 10 (Waltman, van Eck, and Noyons 2010). An overlay visualisation coloured terms by average publication year. Co-citation analysis (Small 1973) mapped the 57 works cited by at least eight corpus documents, after 41 cited-reference labels had been corrected or merged (for example, works mis-attributed to book reviewers and duplicate editions; Supplementary File S4).

To test directly whether simulation-style debriefing has entered L2 research, all 88 corpus records containing debrief *were read and classified by how they used the term (*Fig. 1*).*

2.5 Eligibility criteria for the systematic review

Studies were eligible if they (a) were empirical (quantitative, qualitative, or mixed); (b) involved learners of an L2 or foreign language, including ESP and professional-purpose learners, at any age; (c) included a designed post-performance reflective phase (debriefing, guided or structured reflection, post-task transcription or self-analysis, self-assessment or self-review of recorded performance, or peer or multi-source feedback processed reflectively) following an oral, interactive, or communicative task; (d) reported learner outcomes, processes, or perceptions related to that phase; and (e) were peer-reviewed journal articles or scholarly book chapters published between 2000 and 2026. Studies were excluded if reflection served only as a research instrument (for example, debriefing interviews or reflective journals used as data), if the reflecting agents were teachers or teacher candidates rather than L2 learners, if the task was exclusively written, or if the report was conceptual, a review, or conference proceedings (Table 1). Reflection on extended experiences rather than on a communicative task, such as post-sojourn debriefing workshops (Erdem Mete 2019), was also outside scope.

2.6 Selection process

An eligibility-oriented candidate pool was drawn from the bibliometric corpus with a stricter rule (debriefing, post-task, guided, structured, or self-reflection terms combined with task terms, excluding teacher-education terms). All 405 candidate titles and abstracts were screened against the criteria. Forty-six reports were then assessed for eligibility in detail, using the full text where open access was available (13 reports) and otherwise the abstract and bibliographic metadata. Two reports were excluded at this stage: a non-empirical literature synthesis and a conference-proceedings paper. Two further records were added by other methods. The audit of all 88 debrief *records identified a companion report of an included study (Rif and Tan 2021), and the authors' reference collection supplied one eligible study not indexed in the corpus (Kandasamy, Tan, and Mohamed Sultan 2022). The audit also led to the reinstatement of two records initially excluded at title/abstract stage. In total, 43 studies reported in 46 reports were included (*Fig. 1*). Screening and coding were conducted by a single reviewer with a complete decision log (Supplementary File S3); this is acknowledged as a limitation in* Section 6.

2.7 Data extraction, appraisal, and synthesis

For each study, the following were extracted: setting and country, learner population, sample size, design, reflective modality, outcome domains, direction of findings, and the main result (Table 4). Reflective modalities were coded inductively into five categories: (A) structured debriefing; (B) post-task transcription or self-analysis of output; (C) guided self-reflection or self-assessment; (D) recording-based self-review; and (E) peer or multi-source feedback processed reflectively. A study could receive more than one code. Outcomes were coded into linguistic accuracy and complexity (CAF), global speaking, fluency, and pronunciation, interactional and pragmatic competence, affect, metacognition and autonomy (including self-assessment calibration), and perceptions.

Methodological quality was appraised with a rubric informed by the Mixed Methods Appraisal Tool (Hong et al. 2018). It rated each study as higher (H), moderate (M), or lower (L) on the basis of sample reporting, comparison or triangulation, outcome measurement, and description of the reflective procedure. Because full texts were available for 14 studies, appraisal for the remaining 29 relied on the information reported in abstracts and metadata, and these ratings should be read as provisional. Heterogeneity in designs and outcome measures precluded meta-analysis. Findings were therefore synthesised narratively by modality, and confidence in each body of evidence was judged by design strength, consistency, and reporting quality (Plonsky and Oswald 2014).

3. BIBLIOMETRIC RESULTS (RQ1)

3.1 Growth and distribution

The corpus comprises 888 documents: 833 articles, 51 book chapters, and 4 reviews. Output grew slowly to 2011 (72 documents for 2000–2011), then steadily through 2012–2018 (160) and 2019–2022 (165), and sharply in 2023–2026 (491 documents, 55.3%), even though 2026 is a partial year (Fig. 2). Publication is highly dispersed: 558 distinct sources appear, 436 of them contributing a single document. The most frequent specialist outlets were *Language Teaching Research* (14), *TESOL Quarterly* (11), *ELT Journal* (9), and *Computer Assisted Language Learning* (9). Among the 612 records with affiliation data, the most productive countries were Indonesia (124), the United States (83), China (45), Japan (29), the United Kingdom (29), and Uzbekistan (28). The included studies follow the same trajectory: 17 of the 43 (39.5%) were published in 2024–2026.

3.2 Thematic structure

The term co-occurrence map (Fig. 3; Table 2) contains 101 terms in three clusters. Cluster 1 (red, 39 terms) concerns task performance and outcome measurement. It is organised around *learner* (459 occurrences), *task* (361), *performance*, *self reflection* (164), *feedback* (163), *confidence*, *fluency*, *accuracy*, *complexity*, *pre-task*, and *control/experimental group*. Cluster 2 (green, 32 terms) concerns teachers, courses, and experiential learning, with *teacher* (313), *classroom*, *experiential learning* (94), *debriefing* (47), *reflective practice*, *teacher education*, and *professional development*. Cluster 3 (blue, 30 terms) concerns reflection as communicative engagement. It includes *reflection* (321), *interaction* (250), *communication*, *awareness*, *role play* (148), *simulation* (112), *engagement*, *communicative competence*, *critical thinking*, *learner autonomy*, *reflective journal*, and *AI*.

The map shows a structural separation. *Debriefing* sits in the teacher and experiential-learning cluster, close to teacher education and professional development, whereas *simulation* and *role play* sit in the engagement cluster, and *self reflection* sits with task-performance outcomes. In other words, the term most closely associated with the post-simulation phase is used mainly in teacher-oriented and experiential discourse, not in research on learners' task performance.

Table 2
Thematic clusters in the term co-occurrence map (n = 888)

Cluster	Terms	Highest-occurrence terms (occurrences)	Interpretation
1 (red)	39	learner (459), task (361), foreign language (222), instruction (210), performance (188), self reflection (164), feedback (163), confidence (149), fluency (96), accuracy (75), complexity (56), control/experimental group (45), pre-task (28)	Task performance and outcome measurement: reflection as an intervention variable evaluated through CAF and affective outcomes
2 (green)	32	teacher (313), classroom (270), course (188), framework (148), technology (139), experiential learning (94), training (90), debriefing (47), teacher education (47), professional development (40), reflective practice (32)	Teachers, courses, and experiential learning: debriefing positioned within teacher-oriented and experiential discourse
3 (blue)	30	reflection (321), activity (296), interaction (250), communication (212), awareness (156), role play (148), engagement (141), simulation (112), communicative competence (88), critical thinking (65), AI (50), learner autonomy (49), reflective journal (38)	Reflection as communicative engagement: simulation and role play linked to awareness, autonomy, and critical thinking

3.3 Thematic evolution

The overlay visualisation (Fig. 4) shows a shift from cognitive and acquisition-oriented terms towards affective, engagement, and technology terms. Among the oldest terms (by average publication year) are *reading* (2018.5), *second language acquisition* (2018.6), *production* (2018.7), and *second language* (2019.4). The most recent include *AI* (2025.3), *integration* (2024.0), *significant improvement* (2024.0), *confidence* (2023.9), *learner autonomy* (2023.9), *communicative competence* (2023.8), *reflective journal* (2023.8), and *critical thinking* (2023.6). Recent work thus increasingly frames reflection as a route to confidence, autonomy, critical thinking, and task engagement (Philp and Duchesne 2016), often in technology- or AI-mediated settings, and evaluates it through pre-test/post-test effectiveness designs.

3.4 Intellectual structure

The co-citation network (Fig. 5; Table 3) contains 57 works and 771 co-citation links (total link strength 1,753) in four clusters. Cluster 1 (red, 20 works) is an experiential and sociocultural base led by Kolb (1984), Vygotsky (1978), and Lantolf (2000), together with communicative competence and qualitative-method references. Cluster 2 (green, 15 works) is a task-performance base centred on Skehan (1998), Bygate, Skehan, and Swain (2001), and research on planning, repetition, and complexity, accuracy, and fluency. Cluster 3 (blue, 14 works) is a task-based pedagogy base built around Ellis (2003) and Willis (1996). Cluster 4 (yellow, 8 works) is an interaction and noticing base comprising Long (1996), Schmidt (1990), focus-on-form research, and corrective feedback.

Two features stand out. First, the task-performance and task-pedagogy bases are closely linked (21.9% of link strength runs between them), whereas the experiential base is the most peripheral: only 21.8% of all link strength involves it, and just 6.6% lies within it. Second, the simulation-debriefing literature (e.g., Lederman 1992; Fanning and Gaba 2007; Crookall 2010) does not appear among works cited eight or more times. Theories of debriefing developed in simulation and health-professions education have therefore not entered the intellectual base of L2 reflection research. The audit of debrief *records confirms this. Of 88 records, only 12 (describing 9 studies) used debriefing as a pedagogical phase after an L2 task. Thirty-nine used the term for research procedures such as peer debriefing or debriefing interviews, 11 concerned teacher education, 7 concerned study-abroad programmes, 9 were conceptual or descriptive, and 10 concerned clinical simulation or other uses (*Fig. 1*).*

Table 3
Intellectual bases identified by co-citation analysis (57 works; total link strength 1,753)

Cluster	Works	Most co-cited works (local citations)	Intellectual base	Link strength within
1 (red)	20	Kolb 1984 (33); Vygotsky 1978 (27); Canale and Swain 1980 (16); Lantolf 2000 (15)	Experiential and sociocultural learning; communicative competence	6.6%
2 (green)	15	Skehan 1998 (30); Bygate, Skehan, and Swain 2001 (23); Robinson 2001 (15); Skehan and Foster 1997 (13)	Cognitive task performance (planning, repetition, CAF)	22.5%
3 (blue)	14	Ellis 2003 (29); Skehan 1996 (20); Ellis 2009 (16); Willis 1996 (13)	Task-based language teaching pedagogy	13.7%
4 (yellow)	8	Long 1996 (22); Schmidt 1990 (19); Doughty and Williams 1998 (13); Lyster and Ranta 1997 (11)	Interaction, noticing, and focus on form	4.6%
Between clusters	–	2–3: 21.9%; 2–4: 8.3%; 3–4: 7.2%; 1–3: 6.2%; 1–2: 5.2%; 1–4: 3.8%	Experiential base most peripheral	52.6% total

Note. Local citations = number of corpus documents citing the work; the full cluster membership is documented in Supplementary File S4. Simulation-debriefing works (e.g., Lederman 1992; Fanning and Gaba 2007; Crookall 2010) did not reach the eight-citation threshold.

4. SYSTEMATIC REVIEW RESULTS (RQ2–RQ3)

4.1 Characteristics of the included studies

The 43 included studies (Table 4) were published between 2000 and 2026, and 27 appeared from 2021 onwards. Most were conducted with university or adult learners (37 studies); six involved primary, kindergarten, or secondary learners. Six studies were set in ESP or professional-purpose courses (management, law, medicine, vocational education, creative industries, and public speaking for English majors). Where the setting was reported, studies came from 19 countries, most often Iran (5), Japan (4), Taiwan (3), and Malaysia, Thailand, Indonesia, and Vietnam (2 each); 11 studies did not report a country in the information available. English was the target language in at least 38 studies; French, Chinese, Spanish, Dutch, and Arabic were also represented. Designs comprised 5 randomised experiments, 11 other controlled or comparison-group studies, 3 single-group pre–post or longitudinal studies, 11 mixed-methods, survey, or correlational studies, 10 qualitative studies, and 3 action-research studies. Sample sizes, where reported, ranged from 15 to 160 learners.

Guided self-reflection or self-assessment was the most frequent modality (15 studies), followed by peer or multi-source feedback with reflection (13), structured debriefing (10), post-task transcription (10), and recording-based self-review (9). Outcomes most often concerned global speaking performance (18 studies) and metacognition or autonomy (18), followed by linguistic accuracy and complexity (14), affect (11), perceptions (8), and interactional or pragmatic competence (5). Figure 6 cross-tabulates modalities and outcomes. Thirty-eight studies reported benefits, three reported mixed results, one reported no added benefit, and one was descriptive. On the appraisal rubric, 5 studies were rated higher, 22 moderate, and 16 lower.

Table 4
Characteristics of the 43 included studies

Study	Setting; learners	N	Design	Mod.	Outcomes	Dir.	App.
Abubaker and Ogdol (2025)	UAE; kindergarten ESL; Young learners	NR	QL: Qualitative (observation, interviews, journals)	A	ORAL, AFF	+	L
Kandasamy, Tan, and Mohamed Sultan (2022)	Malaysia; online primary ESL; Year 5 pupils; 11 teachers	50 + 11	CT: Mixed methods; two-group comparison	A	ORAL, PER	+	M
Muranoi (2000)	Japan; university EFL; Adult EFL	91	CT: Quasi-experimental (pre, 2 post)	A	LIN	+	H
Piurko (2026)	Lithuania; university ESP; Undergraduates (creative programmes)	68	MX: Post-intervention survey	A	AFF, INT, PER	+	L
Rif, Tan, and Khairuddin (2023); Rif and Tan (2021)	Malaysia; online primary ESL; Young learners (via teachers)	2 teachers	QL: Qualitative case study (interviews, lesson observation)	A	ORAL, MET	+	M
Velásquez-Hoyos (2021)	Colombia; university EFL; Pre-intermediate adults	NR	AR: Action research	A	ORAL	+	L
Wang (2026)	China; vocational ESP; Vocational undergraduates	48 h recordings	QL: Qualitative case study (recordings, stimulated recall, artefacts)	A	LIN, AFF, MET	mix	M
Yu et al. (2025)	Netherlands; medical Dutch L2; International medical students	16	MX: Mixed methods (survey, interviews)	A, E	MET, PER	+	M
deHaan (2012); deHaan and Johnson	Japan; university EFL; Adult EFL	NR	QL: Design-based intervention; transcript comparison, surveys, interviews	A, B, D	LIN, ORAL, AFF	+	M

Note. N = learners unless stated; NR = not reported in the information available. Design: RCT = randomised experiment; CT = other controlled/comparison design; SG = single-group pre–post or longitudinal; MX = mixed/survey/correlational; QL = qualitative; AR = action research. Mod. (modality): A = structured debriefing; B = post-task transcription; C = guided self-reflection/self-assessment; D = recording-based self-review; E = peer/multi-source feedback with reflection. Outcomes: LIN = accuracy/CAF; ORAL = global speaking/fluency/pronunciation; INT = interactional/pragmatic; AFF = affective; MET = metacognitive/autonomy/calibration; PER = perceptions. Dir. (direction): + = benefit; mix = mixed; 0 = no added benefit; desc = descriptive. App. (appraisal): H/M/L = higher/moderate/lower. Full-text basis for 14 studies (Supplementary File S6).

Study	Setting; learners	N	Design	Mod.	Outcomes	Dir.	App.
(2012a, 2012b)							
van Compernolle and Ballesteros Soria (2020)	NR; L2 classroom; Adult L2	3 groups	QL: Qualitative microgenetic interaction analysis	A	INT	+	M
Ganji, Jafari, and Asgari (2020)	Iran; university EFL; Low and high proficiency	4 groups	CT: Quasi-experimental (12 weeks)	B	ORAL, MET	+	M
Hassanzadeh-Taleshi, Yaqubi, and Bozorgian (2023)	Iran; EFL; Intermediate	38	RCT: Randomised two-group experiment	B	LIN	0	H
Hsu (2019)	Taiwan; university EFL; Adult EFL	39	RCT: Randomised three-group experiment	B	LIN	+	H
Huang (2008)	Taiwan; university EFL; Adult EFL dyads	NR	QL: Descriptive/qualitative	B	MET	+	L
Jafari, Ketabi, and Tavakoli (2016)	Iran; adult EFL; Intermediate and advanced	4 classes	CT: Quasi-experimental (20 weeks)	B	LIN	+	M
Khramchenko (2026)	Russia; MA ESP (management); Postgraduates	32	MX: Mixed methods (corpus + reflective essays)	B, C	INT, MET	+	M
Nguyen et al. (2025)	Vietnam; university EFL; English majors	27	RCT: Randomised three-group experiment	B, E	LIN	+	M
Saeedi (2023)	Iran; EFL; Adult EFL	60	CT: Four-condition experiment	B	LIN	mix	M
Skehan et al. (2012)	NR; Adult EFL	NR	CT: Experimental (three studies; study 3 on transcription)	B	LIN	+	M
Abdelhalim and Alsalem (2019)	Saudi Arabia; EFL and AFL; Adult women	30	SG: Pre–post, two experimental groups	C, D	LIN, AFF	+	L
Alfianti (2022)	Indonesia; university	34	RCT: Randomised posttest-only control	C, D	ORAL	+	M

Note. N = learners unless stated; NR = not reported in the information available. Design: RCT = randomised experiment; CT = other controlled/comparison design; SG = single-group pre–post or longitudinal; MX = mixed/survey/correlational; QL = qualitative; AR = action research. Mod. (modality): A = structured debriefing; B = post-task transcription; C = guided self-reflection/self-assessment; D = recording-based self-review; E = peer/multi-source feedback with reflection. Outcomes: LIN = accuracy/CAF; ORAL = global speaking/fluency/pronunciation; INT = interactional/pragmatic; AFF = affective; MET = metacognitive/autonomy/calibration; PER = perceptions. Dir. (direction): + = benefit; mix = mixed; 0 = no added benefit; desc = descriptive. App. (appraisal): H/M/L = higher/moderate/lower. Full-text basis for 14 studies (Supplementary File S6).

Study	Setting; learners	N	Design	Mod.	Outcomes	Dir.	App.
	public speaking; First-year EFL		design				
Almashour, Aldamen, and Jarrah (2026)	Jordan; EFL; University learners	160	SG: Longitudinal secondary analysis (mixed-effects) + reflections	C	ORAL, MET	+	H
Chovancová (2024)	NR; legal ESP; University law students	NR	QL: Qualitative (student comments)	C	MET, INT	+	L
Hsu and Chen (2025)	NR; EFL beyond the classroom; University learners	1 class (2 groups)	CT: Quasi-experimental	C	ORAL	+	M
Kamolova (2026)	NR; pedagogical university EFL; Intermediate learners	NR	CT: Quasi-experimental mixed methods	C	ORAL, INT, AFF	+	L
Khezrlou (2021)	Iran; higher education; Adult EFL	31	RCT: Randomised two-group experiment	C	LIN	+	M
Konchiab and Munpanya (2021)	Thailand; university EFL; Third-year undergraduates	47	MX: Mixed methods (teacher vs self-assessment)	C	MET, PER	desc	M
Lasso de la Vega Moreno (2025)	Costa Rica; university EFL; Upper-intermediate	15	QL: Qualitative classroom study	C, E	MET, PER	+	L
Lee and Mori (2021)	Japan; university L2 speaking; Undergraduates	25 (+ 11 interviews)	MX: Explanatory mixed methods	C, E	MET	+	M
Meritan (2022)	NR; French L2; University learners	72	CT: Quasi-experimental, three semesters	C	ORAL	+	H
Song (2017)	Korea; high school EFL; Adolescents	30	CT: Quasi-experimental + retrospective interviews	C	LIN, MET	+	M

Note. N = learners unless stated; NR = not reported in the information available. Design: RCT = randomised experiment; CT = other controlled/comparison design; SG = single-group pre–post or longitudinal; MX = mixed/survey/correlational; QL = qualitative; AR = action research. Mod. (modality): A = structured debriefing; B = post-task transcription; C = guided self-reflection/self-assessment; D = recording-based self-review; E = peer/multi-source feedback with reflection. Outcomes: LIN = accuracy/CAF; ORAL = global speaking/fluency/pronunciation; INT = interactional/pragmatic; AFF = affective; MET = metacognitive/autonomy/calibration; PER = perceptions. Dir. (direction): + = benefit; mix = mixed; 0 = no added benefit; desc = descriptive. App. (appraisal): H/M/L = higher/moderate/lower. Full-text basis for 14 studies (Supplementary File S6).

Study	Setting; learners	N	Design	Mod.	Outcomes	Dir.	App.
Vo and Le (2026)	Vietnam; university EFL; Undergraduates	26	AR: Action research (two cycles)	C, E	ORAL, MET	+	L
Winke (2014)	NR; Chinese FL; Advanced university learners	17	MX: Mixed methods (self vs expert ratings)	C, E	MET	+	M
Badianti (2026)	Indonesia; university EFL; Fourth-semester students	32	AR: Classroom action research (two cycles)	D	AFF, ORAL	+	L
El Mortaji (2018)	Morocco; university public speaking; First-year students	20	MX: Mixed methods (60 speeches, surveys, essays)	D	ORAL, AFF	+	L
Qureshi, Roseberry, and Qureshi (2020)	Korea and Pakistan; university EFL; Sophomores	85	MX: Mixed methods	D, E	ORAL, PER	+	L
Shih (2010)	Taiwan; ESP public speaking; Seniors majoring in English	44	MX: Mixed methods	D, E	PER	+	L
Sulistyo et al. (2026)	NR; university EFL; Second-year undergraduates	2 intact groups	CT: Quasi-experimental convergent mixed methods	D, E	ORAL, LIN, AFF	mix	M
Villamizar and Mejía (2019)	NR; Spanish FL; University learners	NR	QL: Qualitative	D	ORAL, AFF, MET	+	L
Kassymova et al. (2025)	NR; secondary FL; Ages 11–15	NR	MX: Survey and observation	E	AFF	+	L
Kireeti, Egitim, and Thomson (2024)	Japan; university EFL presentations; First/second-year students	17 + 87	MX: Sequential explanatory mixed methods	E	MET, PER	+	M

Note. N = learners unless stated; NR = not reported in the information available. Design: RCT = randomised experiment; CT = other controlled/comparison design; SG = single-group pre–post or longitudinal; MX = mixed/survey/correlational; QL = qualitative; AR = action research. Mod. (modality): A = structured debriefing; B = post-task transcription; C = guided self-reflection/self-assessment; D = recording-based self-review; E = peer/multi-source feedback with reflection. Outcomes: LIN = accuracy/CAF; ORAL = global speaking/fluency/pronunciation; INT = interactional/pragmatic; AFF = affective; MET = metacognitive/autonomy/calibration; PER = perceptions. Dir. (direction): + = benefit; mix = mixed; 0 = no added benefit; desc = descriptive. App. (appraisal): H/M/L = higher/moderate/lower. Full-text basis for 14 studies (Supplementary File S6).

Study	Setting; learners	N	Design	Mod.	Outcomes	Dir.	App.
Tsai and Piamsai (2025)	Thailand; secondary EFL; Grade 12	31	SG: Mixed methods (single group)	E	LIN, ORAL, MET	+	M
Yin (2017)	NR; EFL presentations; University learners	17	QL: Qualitative (reflective journals)	E	MET	+	L

Note. N = learners unless stated; NR = not reported in the information available. Design: RCT = randomised experiment; CT = other controlled/comparison design; SG = single-group pre–post or longitudinal; MX = mixed/survey/correlational; QL = qualitative; AR = action research. Mod. (modality): A = structured debriefing; B = post-task transcription; C = guided self-reflection/self-assessment; D = recording-based self-review; E = peer/multi-source feedback with reflection. Outcomes: LIN = accuracy/CAF; ORAL = global speaking/fluency/pronunciation; INT = interactional/pragmatic; AFF = affective; MET = metacognitive/autonomy/calibration; PER = perceptions. Dir. (direction): + = benefit; mix = mixed; 0 = no added benefit; desc = descriptive. App. (appraisal): H/M/L = higher/moderate/lower. Full-text basis for 14 studies (Supplementary File S6).

4.2 Structured debriefing (Modality A; 10 studies)

Structured debriefing was studied in simulation-based, scenario-based, and strategic-interaction tasks. The earliest and most rigorous study is Muranoi's (2000) quasi-experiment with 91 Japanese EFL learners. It showed that interaction enhancement followed by a form-focused debriefing produced greater gains in article use than the same interaction followed by a meaning-focused debriefing. The content of the debriefing therefore mattered, not merely its presence. Kandasamy, Tan, and Mohamed Sultan (2022) compared online primary ESL classes with and without an exit-ticket debriefing and reported higher test scores in the debriefing group. In the same research programme, teachers reported gains in learners' critical thinking and proficiency, but also learner reticence and difficulty in formulating debriefing questions (Rif and Tan 2021; Rif, Tan, and Khairuddin 2023).

Several studies documented how debriefing works in interaction. In van Compernolle and Ballesteros Soria's (2020) dynamic strategic interaction scenarios, a whole-class debriefing after one group's performance made turn allocation an explicit focus, and later groups took this up in their own performances, a collective form of feedforward. deHaan's (2012) wiki-mediated strategic-interaction cycles combined video recording, self-transcription, self-correction, and group debriefing, and transcript comparisons showed syntactic, pragmatic, lexical, and fluency development across cycles (see also deHaan and Johnson 2012a, 2012b). Wang (2026) showed that translanguaging-supported metalinguistic reflection during debriefing in a vocational ESP curriculum supported self-efficacy and complexity, but warned of 'comfort traps' when accountability tools were absent. In a medical Dutch course built on briefing–simulated consultation–debriefing cycles, students trusted external feedback more than their own self-reflections yet judged both equally useful (Yu et al. 2025). The remaining studies reported confidence, participation, fluency, and discipline-specific vocabulary gains associated with debriefed simulations, panels, and rehearsal–performance–debriefing workshops, but relied on observation, surveys, or action research (Abubaker and Ogdol 2025; Piurko 2026; Velásquez-Hoyos 2021).

Confidence in this body of evidence is low. Only two studies included a comparison condition, outcome measures were rarely validated, and none compared debriefing structures of the kind developed in simulation education (e.g., reaction–analysis–summary or advocacy–inquiry; Rudolph et al. 2006; Sawyer et al. 2016).

4.3 Post-task transcription and self-analysis (Modality B; 10 studies)

Transcription-based reflection, in which learners transcribe and analyse their own recorded output, is the most experimentally developed modality, building on Lynch's (2001, 2007) proposals. Hsu's (2019) randomised three-group study found that task repetition with post-task transcribing produced more accurate repeated performances than repetition alone, and that the accuracy advantage transferred to a new task. Nguyen et al. (2025) found that teacher-corrected transcribing combined with repetition increased complexity and fluency on repeated tasks and accuracy and complexity on a new task. Saeedi (2023) showed that the mere anticipation of post-task transcription raised accuracy, maximally when combined with pre-task planning but at the cost of complexity, consistent with Foster and Skehan (2013). Skehan et al. (2012) likewise associated post-task transcription conditions with fewer errors. Two quasi-experiments with university learners in Iran found that transcribing recorded group discussions, followed by self- and peer correction and reformulation, improved grammatical accuracy over 20 weeks (Jafari, Ketabi, and Tavakoli 2016) and oral proficiency and autonomy over 12 weeks (Ganji, Jafari, and Asgari 2020). Descriptive studies showed that cooperative transcription directed learners' attention to their own errors (Huang 2008) and that transcripts of AI-persona negotiations supported self-observation of pragmatic choices in ESP (Khramchenko 2026).

The evidence is not uniformly positive. Hassanzadeh-Taleshi, Yaqubi, and Bozorgian's (2023) randomised study found that immediate post-task transcribing added nothing to task repetition on any CAF measure, in either immediate or delayed repetitions. A plausible reading of the evidence as a whole is that transcription works when it is followed by analysis and correction, whether by the self, peers, or teacher. Transcription alone is not enough. Confidence in accuracy effects is moderate, and evidence for other outcomes is limited.

4.4 Guided self-reflection and self-assessment (Modality C; 15 studies)

Guided reflection between repeated performances has produced consistent accuracy effects. Khezrlou (2021) randomly assigned learners to task repetition with or without a Kolb-based reflection questionnaire and found that reflection improved past-tense accuracy on both repeated and new tasks. Song (2017) reported that self-reflection during repetition raised awareness of target forms and improved accuracy among Korean high-school learners. Over three semesters, Meritan (2022) found that self-reflection combined with explicit pronunciation instruction outperformed instruction alone on the intelligibility of French vowels. Hsu and Chen (2025) found that adding self-reflection forms to practice with intelligent personal assistants improved listening and speaking beyond a practice-only group.

A second line of work examined self-assessment and its calibration. Winke (2014) showed that advanced learners of Chinese, after self-rating their role-play performances and discussing those ratings against expert ratings in a reflective lesson, were realistic judges of their accuracy, fluency, complexity, and tones. In the largest included study, Almashour, Aldamen, and Jarrah (2026) found among 160 Jordanian EFL learners that more accurate initial calibration between self-assessed and listener-rated comprehensibility predicted steeper comprehensibility growth, whereas overconfidence predicted flatter trajectories. Post-task reflections showed that accurate self-assessors used listener-oriented criteria. In contrast, Konchiab and Munpanya (2021) found that Thai undergraduates rated their language features lower than their content, which suggests that learners' self-assessment criteria differ from teachers'. Self-assessment of recorded speaking improved performance in a posttest-only controlled study (Alfianti 2022). Other studies linked reflective practices to self-directed learning competencies (Lee and Mori 2021), deeper reflection when theory was supplied after a legal ESP role play (Chovancová 2024), and stronger engagement with feedback when AI-assisted prompts guided error analysis (Lasso de la Vega Moreno 2025). Confidence is moderate for accuracy effects of reflection between repetitions and low to moderate for the other outcomes.

4.5 Recording-based self-review (Modality D; 9 studies)

Video or audio recording gives learners a durable record of otherwise ephemeral speech. Studies of public speaking and presentation courses reported that videotaping combined with self-reflection improved perceived content, delivery, organisation, and confidence (El Mortaji 2018), supported peer feedback and self-reflection across successive speeches (Qureshi, Roseberry, and Qureshi 2020), and was associated with learning satisfaction in an ESP video-blog course (Shih 2010). Recording-based reflection was also associated with reduced speaking anxiety (Abdelhalim and Alsalem 2019; Badianti 2026; Villamizar and Mejía 2019). The only controlled study in this group gave a mixed picture: video projects with dialogic reflection improved overall speaking and fluency but not complexity or accuracy (Sulistyo et al. 2026). The pattern suggests that self-review supports affective and global performance outcomes more readily than linguistic precision, unless it is paired with focused analysis of the kind used in Modality B. Most studies were single-group, perception-based, or action research, so confidence is low.

4.6 Peer and multi-source feedback with reflection (Modality E; 13 studies)

Many studies located reflection within a social feedback process. Multi-source feedback from teachers and peers over repeated presentations deepened learners' self-reflection (Yin 2017). Peer evaluation was associated with learner autonomy satisfaction (Kireeti, Egitim, and Thomson 2024). Peer assessment combined with non-scripted role plays improved CAF and self-evaluation among Thai secondary learners (Tsai and Piamsai 2025). Adding rubrics, peer feedback, and self-reflection forms to e-portfolios improved speaking and learner responsibility (Vo and Le 2026). In formative oral assessment, feedback, peer assessment, and self-reflection were associated with reduced distress (Kassymova et al. 2025). External comparison also appeared as the mechanism that made self-reflection productive in studies already discussed (Winke 2014; Yu et al. 2025; Nguyen et al. 2025). Confidence is low because most studies lacked comparison conditions, but the convergence on external comparison as an enabling condition is noteworthy.

Table 5
Synthesis of evidence by reflective modality

Modality	Studies	Designs	Direction	Principal findings	Confidence
A. Structured debriefing	10	AR 1, CT 2, MX 2, QL 5	+ 9, mixed 1	Form-focused debriefing > meaning-focused debriefing (one quasi-experiment); collective uptake of debriefed interactional features; gains in confidence, participation, fluency, and terminology mostly from qualitative and survey evidence	Low
B. Post-task transcription	10	CT 4, MX 1, QL 2, RCT 3	+ 8, mixed 1, null 1	Accuracy gains on repeated and new tasks when transcription is followed by correction; anticipation alone raises accuracy; one randomised null result for transcription without further analysis; accuracy–complexity trade-off	Moderate (accuracy)
C. Guided self-reflection / self-assessment	15	AR 1, CT 4, MX 4, QL 2, RCT 2, SG 2	+ 14, descriptive 1	Reflection between repetitions improves accuracy; calibration of self-assessment predicts comprehensibility growth; self-assessment supports autonomy and engagement with feedback	Moderate (accuracy); low to moderate (other)
D. Recording-based self-review	9	AR 1, CT 1, MX 3, QL 2, RCT 1, SG 1	+ 8, mixed 1	Reduced anxiety and greater confidence; perceived gains in delivery and content; overall speaking/fluency gains without accuracy or complexity gains in the one controlled study	Low
E. Peer / multi-source feedback with reflection	13	AR 1, CT 1, MX 7, QL 2, RCT 1, SG 1	+ 12, mixed 1	External comparison deepens self-reflection and autonomy; peer assessment with role play linked to CAF and self-evaluation gains	Low

Note. Studies may use more than one modality. Design codes as in Table 4. Confidence judged from design strength, consistency, and reporting quality.

4.7 Methodological quality

Table 6 summarises the appraisal. The higher-rated studies (Muranoi 2000; Hsu 2019; Hassanzadeh-Taleshi, Yaqubi, and Bozorgian 2023; Meritan 2022; Almashour, Aldamen, and Jarrah 2026) share comparison conditions or longitudinal modelling, adequate samples, and objective outcome measures. Recurrent weaknesses were small or unreported samples, absent comparison groups, reliance on learner perceptions, outcome measures without reported reliability, and insufficient description of the reflective procedure itself: who led it, how long it lasted, which prompts were used, and whether it referred to explicit criteria. Effect sizes were rarely reported in the information available, which prevented quantitative synthesis.

Table 6
Methodological appraisal summary (MMAT-informed rubric)

Rating	Studies	Typical characteristics
Higher (H)	5	Comparison condition or longitudinal modelling, adequate sample, objective outcome measures: Muranoi 2000; Hsu 2019; Hassanzadeh-Taleshi, Yaqubi, and Bozorgian 2023; Meritan 2022; Almashour, Aldamen, and Jarrah 2026
Moderate (M)	22	Comparison or triangulation present but small samples, limited reliability reporting, or incomplete procedural detail
Lower (L)	16	Single-group, perception-only, or action-research designs; unreported samples; unvalidated outcomes; reflective procedure insufficiently described
Cross-cutting gaps	43	Effect sizes rarely reported; delayed post-tests rare; reflective procedure (leader, duration, prompts, criteria) seldom specified; interactional and professional outcomes rarely measured

Note. Appraisal used full texts for 14 studies and abstracts/metadata for 29; ratings for the latter are provisional.

5. DISCUSSION (RQ4)

5.1 A fragmented field with a missing theoretical bridge

The bibliometric and systematic evidence converge on one conclusion. Post performance activities of L2 communicative tasks have been adopted and are studied extensively, although studied in bits and pieces. The term map situates debriefing in teacher-oriented and experiential discourse, and separates it from research on task-performance outcomes. The co-citation map indicates that L2 reflection research is based on the tradition of experiential, sociocultural, task-based and noticing traditions, but not on the debriefing theory developed in the field of simulation and health-professions education. There are only 12 records in a total of 88 records indicating a pedagogical stage following an L2 task, using the term debriefing.Only 12 of 88 records mention a pedagogical stage following the L2 task, and describe it as a debriefing. L2 researchers therefore use simulations without the debriefing science that normally accompanies them, while debriefing scholars have not engaged with L2 learning. The review's first contribution is to document this disconnect and to bring the two literatures together.

5.2 What the evidence supports

Meta-analytic evidence shows that task-based programmes are effective overall (Bryfonski and McKay 2019), but it does not isolate the post-task phase. In the recent review below, the best evidence is for linguistic correctness. In the case of analysing learners' own work in transcription and correction (Hsu 2019; Nguyen et al. 2025; Jafari, Ketabi, and Tavakoli 2016), and guided reflection between repetitions (Khezrlou 2021; Song 2017), accuracy is increased on repeated and sometimes new tasks. This is in line with the noticing hypothesis (Schmidt 1990) and evidence that expected post task activities focus attention on form (Foster and Skehan 2013). There are two qualifications from the evidence. First, the reflective activity should not be just about exposure to one's own output, but should be an analysis and correction of the output (Hassanzadeh-Taleshi, Yaqubi, and Bozorgian 2023). Second, increases in accuracy might be at the cost of complexity (Saeedi 2023), similar to the trade-off effects found in task-based studies (Skehan 1998).

Muranoi's (2000) initial research comparing form-focused debriefing to meaning-focused debriefing has never been replicated in over 20 years and was not a direct comparison of the content of debriefing. Recording-based and peer-

supported reflection is consistently associated with lower anxiety and greater confidence and autonomy. These are valuable outcomes in ESP and presentation courses, but most of the supporting designs cannot separate the effect of reflection from that of repeated practice. Evidence on calibration (Winke 2014; Almashour, Aldamen, and Jarrah 2026) suggests a further mechanism: reflection works best when learners' internal standards are aligned with external ones. This agrees with research on self-assessment and self-regulated learning in general education (Andrade 2019; Panadero, Jonsson, and Botella 2017; Nicol and Macfarlane-Dick 2006).

5.3 Reconciling cognitive and sociocultural accounts of reflection

The included studies rest on two theoretical accounts that are rarely brought together. Experimental studies of transcription and guided reflection adopt a cognitive account, in which reflection frees attentional resources and directs them to form, so that learners notice gaps and restructure their output (Schmidt 1990; Skehan 1998). Qualitative studies of debriefing and peer feedback adopt a sociocultural account, in which reflection is a mediated, dialogic activity that makes new interactional resources available to the group (Vygotsky 1978; Lantolf 2000). The co-citation map reproduces this division: the task-performance and noticing bases are densely interlinked, whereas the experiential and sociocultural base is peripheral. The evidence suggests that the two accounts describe complementary mechanisms rather than competing ones. In van Compernolle and Ballesteros Soria (2020), for example, a public debriefing turned one group's performance into a shared object of attention, a sociocultural process whose effect was a cognitive one, namely increased attention to turn-allocation devices. Conversely, the transcription studies that worked best added a social layer of peer or teacher correction to individual noticing (Jafari, Ketabi, and Tavakoli 2016; Nguyen et al. 2025). Future studies could test this integration directly by comparing individual and dialogic reflection on the same transcribed performance.

5.4 What L2 research can borrow from simulation debriefing

Simulation education offers three resources that the L2 literature has not yet used. The first is the phase structure. Debriefing frameworks include a first, emotional, stage of debriefing followed by a stage of analysis and summary (Fanning and Gaba 2007; Eppich and Cheng 2015). For L2 learners, separation is a factor that is of concern in the studies included as it was a recurrent result in the studies on L2 learners' anxiety after performance in a foreign language (Abdelhalim and Alsalem 2019; Badianti 2026). The second is a position for the facilitator. The advocacy–inquiry approach is an approach that integrates an explicit judgement regarding the performance with a genuine question regarding the reasoning of the learner (Rudolph et al. 2006) which could support teachers to tackle the difficulty in formulating debriefing questions mentioned by Rif, Tan, and Khairuddin (2023). The third is an evidence base on blended and learner-led formats (Sawyer et al. 2016) that could inform how teacher-led, peer-led, and self-led reflection are combined in large L2 classes. Adapting these resources to L2 targets, by adding language-, interaction-, and task-focused analysis to the analysis phase, is the most direct route to closing the gap identified in this review.

5.5 A post-performance reflective cycle for L2 and ESP pedagogy

Figure 7 integrates the findings into a five-phase post-performance reflective cycle. In (1) *communicative performance*, learners carry out a role play, simulation, presentation, or negotiation. In (2) *evidence capture*, the performance is preserved through recording, transcription, observation notes, or ratings, because reflection on ephemeral speech is otherwise limited to memory (Lynch 2001; Qureshi, Roseberry, and Qureshi 2020). Structured reflection (3): A guided sequence from reaction to language, interaction and task description to analysis. This takes a phased debriefing approach to simulation education (Fanning and Gaba 2007; Rudolph et al. 2006) and applies it to L2 targets. Calibration (4) is the condition that teacher, peer, expert, and/or AI feedback help learners compare

their judgment with to make self-reflection productive (Winke 2014; Yu et al. 2025; Almashour, Aldamen, and Jarrah 2026). In (5) *feedforward*, learners turn the analysis into goals and strategies for a repeated or subsequent performance (Hattie and Timperley 2007; van Compernolle and Ballesteros Soria 2020). The model specifies four moderating conditions supported by the synthesis: guidance quality, external comparison, opportunity to re-perform, and learner factors such as age, proficiency, anxiety, and calibration. It is offered as a testable framework, not as an established causal model.

5.6 Implications for ESP practice

The synthesis helps ESP teachers who incorporate simulations and role plays in their lessons with four practical principles. The first step is to plan the reflective phase as carefully as the task and include explicit time, prompts, and criteria for the reflection; the least structured form of 'how did it go' is the least well documented. Second, record (or transcribe) the evidence so that reflection can focus on what was said. Third, as noted above, provide an outside standard for reflection by providing rubric descriptors, peer, teacher or expert ratings, or models to use as a point of comparison. Finally, close the loop and ask for an improvement plan for a follow-up or repeated task. The number of ESP studies in the review does not diminish, and there are some studies that address professional aspects of communication not adequately captured by general L2 studies, such as pragmatic negotiation, pragmatic negotiation strategies in attorney–client meetings, and doctor–patient communication (Khramchenko 2026; Chovancová 2024; Yu et al. 2025). If reflection is specifically evaluated, some grade descriptors for reflective depth in the field of professional education can be adapted (Wald et al. 2012).

6. LIMITATIONS

These conclusions are subject to a number of limitations. For the first time only one open bibliographic source was used. OpenAlex is wide but has noiser metadata than curated databases and might not have studies indexed in Web of Science or Scopus or written in languages other than English. Second, screening and coding was done by one coder. A full decision log is available but inter-rater agreement was not calculated and a random subset of at least one copy of the decision log should be performed separately before the journal submission. Third, full texts were obtainable for just 14 of the 43 studies, and extraction and appraisal for the other studies was based on abstracts and metadata, with some ratings potentially changing once studies were read in full. The fourth is that conference like and less visible outlets expand coverage, but also bring in some variability in quality. Fifth, due to the lack of effect sizes reported and the variation in outcome measures, a meta-analysis was not possible, and publication bias might have influenced the narrative synthesis: 38 of 43 studies reported benefits. Lastly, thresholds and cleaning operations in the bibliometric maps and co-citation maps depend on a series of decisions that are reported to allow sensitivity analysis.

7. RESEARCH AGENDA AND CONCLUSION

The gaps identified are the basis of a research agenda presented in Table 7. The most immediate needs include the development of controlled comparisons of debriefing formats and content in L2 simulations (based on Muranoi, 2000) and on models of simulation-education debriefing. Other needs include measuring interactional and professional outcomes, not just linguistic outcomes, separating the effect of reflection from repetition, employing delayed post-tests, conducting research with young learners and reflecting on the ESP workplace, and providing clear reporting of reflective procedures and effect sizes (Plonsky and Oswald 2014).

Table 7
Research agenda

Gap identified	Evidence	Recommended research
Debriefing formats untested in L2	Only one controlled comparison of debriefing content (Muranoi 2000); simulation-debriefing theory absent from co-citation base	Controlled comparisons of structured debriefing models (e.g., reaction–analysis–summary, advocacy–inquiry) adapted to L2 targets
Reflection confounded with repetition	Most designs combine reflection with re-performance	Factorial designs crossing reflection × repetition with delayed post-tests
Narrow outcomes	Interactional/pragmatic outcomes in 5 studies; professional ESP outcomes rare	Measure interactional competence and domain-specific task accomplishment alongside CAF
Calibration mechanism	Two studies link self–external calibration to development	Test calibration feedback as a mediator; use rubric-based self-assessment training
Contexts and learners	Few young-learner, workplace, and ESP studies	Classroom and workplace ESP simulations; age- and proficiency-sensitive designs
Technology and AI	AI-mediated reflection emerging since 2024	Compare human-, peer-, and AI-guided debriefing; examine over-reliance and accuracy of AI feedback
Reporting quality	Effect sizes and procedures seldom reported	Report reflective procedures in detail and standardised effect sizes (Plonsky and Oswald 2014)

Concluding the review, there were 888 publications mapped and 43 empirical studies synthesized in what L2 learners do after L2 communicative performance. Analysis of their own work (usually recorded, directed and referenced to external standards) is linked to improvements in accuracy, confidence and independence. Structured debriefing, which is a pedagogical tool of simulation to establish its effectiveness, is however under-theorised and under-tested in L2 and ESP research. Linking L2 task research to L2 debriefing of science, and to controlled classroom studies of the proposed reflective cycle, provides a clear path for future research.

Declarations

Ethics approval and consent to participate.

Not applicable: this study is a review of published literature and involved no human participants.

Consent for publication.

Not applicable.

Funding.

Not applicable.

Availability of data and materials.

The OpenAlex query results, screening code, thesaurus, VOSviewer map, network and corpus files, the reference-disambiguation audit, the debriefing-term audit, the screening decision log, and the study-level extraction sheet are provided as Supplementary Files S1–S6.

Competing interests. The authors declare no competing interests.

References

1. Abdelhalim SM, Alsalem AS (2019) Developing oral linguistic performance and alleviating speaking anxiety of EFL and AFL learners: A proposed strategy. Int J Educ 11(3):107–135. https://doi.org/10.5296/ije.v11i3.15430
2. Abubaker A, Ogdol R (2025) Simulation-based instruction to promote kindergarteners' speaking skills. Int J Res Educ Sci 11(4):871–881. https://doi.org/10.46328/ijres.1319
3. Alfianti M (2022) The implementation of self-assessment in EFL student's speaking performance. J Engl Lang Teach Literature (JELTL) 5(2):103–116. https://doi.org/10.47080/jeltl.v5i2.2043
4. Almashour M, Aldamen H, Jarrah M (2026) Initial self-assessment calibration and post-baseline L2 speech comprehensibility trajectories: A secondary longitudinal analysis. Social Sci Humanit Open 14. https://doi.org/10.1016/j.ssaho.2026.103167. Article 103167
5. Andrade HL (2019) A critical review of research on student self-assessment. Front Educ 4. : Article 87. https://doi.org/10.3389/feduc.2019.00087
6. Badianti A (2026) Reducing students' speaking anxiety through video self-recording in speaking class at UIN Sayyid Ali Rahmatullah Tulungagung. Global Educ J 4(1):643–652. https://doi.org/10.59525/gej.1920
7. Bryfonski L, McKay TH (2019) TBLT implementation and evaluation: A meta-analysis. Lang Teach Res 23(5):603–632. https://doi.org/10.1177/1362168817744389
8. Bygate M, Skehan P, Swain M (eds) (2001) Researching pedagogic tasks: Second language learning, teaching and testing. Longman, Harlow
9. Canale M, Swain M (1980) Theoretical bases of communicative approaches to second language teaching and testing. Appl Linguist 1(1):1–47. https://doi.org/10.1093/applin/I.1.1
10. Chovancová B (2024) Training legal skills in the ESP classroom: Mediation activities and student self-reflection. HERMES - J Lang Communication Bus 64:137–148. https://doi.org/10.7146/hjlcb.vi64.147478
11. Crookall D (2010) Serious games, debriefing, and simulation/gaming as a discipline. Simul Gaming 41(6):898–920. https://doi.org/10.1177/1046878110390784
12. Culbert JH, Hobert A, Jahn N, Haupka N, Schmidt M, Donner P, Mayr P (2025) Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. Scientometrics 130(4):2475–2492. https://doi.org/10.1007/s11192-025-05293-3
13. deHaan J (2012) Wiki and digital video use in strategic interaction-based experiential EFL learning. CALICO J 29(2):249–268. https://doi.org/10.11139/cj.29.2.249-268
14. deHaan J, Johnson NH (2012a) Second language strategic interactions using emerging technologies and experiential learning. In Educational stages and interactive learning: From kindergarten to workplace training, 306–330. Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-4666-0137-6.ch018
15. deHaan J, Johnson NH (2012b) Enhancing the scenario: Emerging technologies and experiential learning in second language instructional design. Int J Learning: Annual Rev 18(4):321–334.

https://doi.org/10.18848/1447-9494/CGP/v18i04/47569
16. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM (2021) How to conduct a bibliometric analysis: An overview and guidelines. J Bus Res 133:285–296. https://doi.org/10.1016/j.jbusres.2021.04.070
17. Doughty C, Williams J (eds) (1998) Focus on form in classroom second language acquisition. Cambridge University Press, Cambridge
18. Ali MR, Peiseniece L, Mikelsone E (2026) Fostering environmental sustainability through green HRM green leadership and work commitment. Discov Sustain. https://doi.org/10.1007/s43621-026-04468-z
19. Ali MR, Peiseniece L, Mikelsone E, Shah ASA, Kadirova Z (2026) Co-Creating Sustainable Innovation From Within: Green HRM Practices Enabling EmployeeGreen Behavior in KIBS Ecosystems Across Pakistan, China, And Uzbekistan. J Manage Practices Humanit Social Sci 10(2):263–277. https://doi.org/10.33152/jmphss-10.2.7
20. Bakhronkulova L, Ali M, Khabirova Z, Azimjon A, Zebo A, Muslima A (2025) June. Artificial Intelligence in Cybersecurity: Threats, Defenses, and Future Directions. In ENVIRONMENT. TECHNOLOGY. RESOURCES. Proceedings of the International Scientific and Practical Conference (Vol. 2, pp. 23–30). https://doi.org/10.17770/etr2025vol2.8592
21. Shah ASA, Ali MR, Vokhidova M, Xusanov C, Musaeva R (2025) Charting the Future: Analyzing Innovative Strategies in Strategic Management for Sustainable Performance Excellence and Business Leadership in the Telecom Industry. Bus Ethics Leadersh 9(4):195–208. https://doi.org/10.61093/bel.9(4).195-208.2025
22. Ali M, Peiseniece L, Mikelsone E (2025) June. Through the green work commitment, a literature review examines the relationship between green HRM and environmental sustainability. In Environment. Technology. Resources. Proceedings of the International Scientific and Practical Conference (Vol. 1, pp. 31–43). https://doi.org/10.17770/etr2025vol1.8635
23. van Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84(2):523–538. https://doi.org/10.1007/s11192-009-0146-3
24. *El Mortaji L (2018) University students' perceptions of videotaping as a teaching tool in a public speaking course. Eur Sci J 14(8):102–118. https://doi.org/10.19044/esj.2018.v14n8p102
25. Ellis R (2003) Task-based language learning and teaching. Oxford University Press, Oxford
26. Ellis R (2009) Task-based language teaching: Sorting out the misunderstandings. Int J Appl Linguistics 19(3):221–246. https://doi.org/10.1111/j.1473-4192.2009.00231.x
27. Eppich W, Cheng A (2015) Promoting excellence and reflective learning in simulation (PEARLS): Development and rationale for a blended approach to health care simulation debriefing. Simul Healthc 10(2):106–115. https://doi.org/10.1097/SIH.0000000000000072
28. Erdem Mete D (2019) Learning from reflection: A case of a post-sojourn debriefing workshop with EFL university students. J Lang Linguistic Stud 15(4):1376–1394. https://doi.org/10.17263/jlls.668487
29. Fanning RM, Gaba DM (2007) The role of debriefing in simulation-based learning. Simul Healthc 2(2):115–125. https://doi.org/10.1097/SIH.0b013e3180315539
30. Foster P, Skehan P (2013) Anticipating a post-task activity: The effects on accuracy, complexity, and fluency of second language performance. Can Mod Lang Rev 69(3):249–273. https://doi.org/10.3138/cmlr.69.3.249
31. *Ganji M, Jafari S, Asgari M (2020) The role of transcribing group discussion task in promoting autonomy and oral proficiency of university EFL learners. J Engl Lang Teach Learn 12(25):131–148. https://doi.org/10.22034/elt.2020.10678

32. *Hassanzadeh-Taleshi M, Yaqubi B, Bozorgian H (2023) The effects of combining task repetition with immediate post-task transcribing on L2 learners' oral narratives. Lang Learn J 51(2):133–144. https://doi.org/10.1080/09571736.2021.1901967
33. Hattie J, Timperley H (2007) The power of feedback. Rev Educ Res 77(1):81–112. https://doi.org/10.3102/003465430298487
34. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, Gagnon M-P et al (2018) Educ Inform 34(4):285–291. https://doi.org/10.3233/EFI-180221. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers
35. Ali M, Peiseniece L, Mikelsone E (2025) June. The mediating role of green work commitment in the relationship between green HR practices, AI elements, and environmental sustainability: a quantitative analysis. In ENVIRONMENT. TECHNOLOGY. RESOURCES. Proceedings of the International Scientific and Practical Conference (Vol. 2, pp. 15–22). https://doi.org/10.17770/etr2025vol2.8578
36. Khan S, Ali M, Akram M, Azizova N, Vokhidova M, Isamukhamedova G (2025) June. Exploring the impact of digital transformation on business model innovation in the retail sector. In Environment. Technology. Resources. Proceedings of the International Scientific and Practical Conference (Vol. 1, pp. 281–288). https://doi.org/10.17770/etr2025vol1.8640
37. Muhammad Ramzan ALI, Aziz MS, Younis M, FOSTERING ENVIRONMENTAL SUSTAINABILITY THROUGH GREEN HUMAN RESOURCE MANAGEMENT, GREEN LEADERSHIP, AND WORK ENGAGEMENT: EXAMINING THE EMERGING ROLE OF ARTIFICIAL INTELLIGENCE (2026). Zenodo. Available at: https://doi.org/10.5281/zenodo.22897287
38. Ali MR, Iqbal SMJ (2022) Examining the mediating influence of green behavior on the impact of green HR practices on environmental sustainability. iRASD J Manage 4(3):479–494
39. Tehmina Rafi MR, Ali Z, Kadirova U, Yousaf, FOREIGN OWNERSHIP, AND OWNERSHIP CONCENTRATION ON CORPORATE FINANCIAL DECISIONS AND FIRM VALUE: EVIDENCE FROM THE TEXTILE SECTOR OF PAKISTAN (2025) IMPACT OF BOARD GENDER DIVERSITY,. Zenodo. Available at: https://doi.org/10.5281/zenodo.17421976
40. Ali M, Kadirova Z, Ismoyilov B (2024) June. Practicing green HRM to achieve environmental sustainability. In Environment. Technology. Resources. Proceedings of the International Scientific and Practical Conference (Vol. 1, pp. 24–33). https://doi.org/10.17770/etr2024vol1.7961
41. *Hsu H-C (2019) The combined effect of task repetition and post-task transcribing on L2 speaking complexity, accuracy, and fluency. Lang Learn J 47(2):172–187. https://doi.org/10.1080/09571736.2016.1255773
42. *Hsu H-L, Chen HH-J (2025) Examining the impact of intelligent personal assistants with self-reflection on language learning beyond the classroom. Comput Assist Lang Learn 1–27 Advance online publication. https://doi.org/10.1080/09588221.2025.2506478
43. *Huang S-C (2008) Raising learner-initiated attention to the formal aspects of their oral production through transcription and reflection. IRAL - Int Rev Appl Linguistics Lang Teach 46(4). https://doi.org/10.1515/IRAL.2008.016
44. *Jafari S, Ketabi S, Tavakoli M (2016) Autonomously noticing incorrect language use: Does it improve EFL learners' grammatical accuracy? J Teach Lang Skills 35(1):85–110. https://doi.org/10.22099/jtls.2016.3684
45. Jones K (1982) Simulations in language teaching. Cambridge University Press, Cambridge
46. *Kamolova HF (2026) Task-based communicative instruction in EFL: Effects on learners' speaking fluency, interactional competence, and willingness to communicate. Mod Am J Linguistics Educ Pedagogy 2(3):6–19

47. *Kandasamy S, Tan KH, Mohamed Sultan FM (2022) The impact of a debriefing strategy in online ESL classrooms. Int J Learn Teach Educational Res 21(3):247–262. https://doi.org/10.26803/ijlter.21.3.13
48. *Kassymova GK, Markhmadova ZK, Tuanaya R, Khoirunnisa A, Fakhriyah F (2025) Formative assessment in enhancing emotional well-being and academic engagement in secondary school students. Asian J Assess Teach Learn 15(2):58–70. https://doi.org/10.37134/ajatel.vol15.2.4.2025
49. Kaygısız S, Akar H (2025) Role-playing as an instructional technique in English as a foreign language and English as a second language settings: A systematic review. Camb J Educ 55(3):441–467. https://doi.org/10.1080/0305764X.2025.2480579
50. *Khezrlou S (2021) Learners' reflective practice between the repeated performances of tasks: Effects on second language development. Dutch J Appl Linguistics 10. https://doi.org/10.51751/dujal9458
51. *Khramchenko DS (2026) Negotiating with scripted AI personas: Task-specific pragmatic performance among International Management MA students. Training. Lang Cult 10(3):32–48. https://doi.org/10.22363/2521-442X-2026-10-3-32-48
52. *Kireeti K, Egitim S, Thomson BJ (2024) Utilizing peer evaluation as a collaborative learning tool: Fostering autonomy satisfaction in English presentation activities. Lang Learn High Educ 14(2):379–400. https://doi.org/10.1515/cercles-2024-0037
53. Knutson S (2003) Experiential learning in second-language classrooms. TESL Can J 20(2):52–64. https://doi.org/10.18806/tesl.v20i2.948
54. Kolb DA (1984) Experiential learning: Experience as the source of learning and development. Prentice-Hall, Englewood Cliffs, NJ
55. *Konchiab S, Munpanya P (2021) Investigating Thai EFL undergraduates' oral presentation performances and experiences, using teacher and student self-assessments. THAITESOL J 34(1):96–117
56. Lantolf JP (ed) (2000) Sociocultural theory and second language learning. Oxford University Press, Oxford
57. de la *Lasso A (2025) Integrating AI-assisted prompts into oral feedback to enhance student agency in the EFL context. Res Pedagogy 15(2):231–242. https://doi.org/10.5937/IstrPed2502231V
58. Lederman LC (1992) Debriefing: Toward a systematic assessment of theory and practice. Simul Gaming 23(2):145–160. https://doi.org/10.1177/1046878192232003
59. *Lee H, Mori C (2021) Reflective practices and self-directed learning competencies in second language university classes. Asia Pac J Educ 41(1):130–151. https://doi.org/10.1080/02188791.2020.1772196
60. Long MH (1996) The role of the linguistic environment in second language acquisition. In Handbook of second language acquisition, edited by W. C. Ritchie and T. K. Bhatia, 413–468. San Diego, CA: Academic Press. https://doi.org/10.1016/B978-012589042-7/50015-3
61. Lynch T (2001) Seeing what they meant: Transcribing as a route to noticing. ELT J 55(2):124–132. https://doi.org/10.1093/elt/55.2.124
62. Lynch T (2007) Learning from the transcripts of an oral communication task. ELT J 61(4):311–320. https://doi.org/10.1093/elt/ccm050
63. MUHAMMAD RA, BAKHTIYOR I, MAHLIYO M, MUHAMMAD FA (2023) Role of Servant Leadership in Work Performance: Identifing the Mediating Relation with Employee Ambidexterity and Moderating with Competitive Orientation. Oeconomia Copernicana 14:181–211
64. AKRAM MF, ALI, M. R., KADIROVA Z, ROLE OF CONSUMER DIGITAL LITERACY IN CONSUMER BEHAVIORAL INTENTIONS (2023a) : MEDIATING EFFECT OF PERCEIVED EASE OF USE, USEFULNESS AND CUSTOMER ENGAGEMENT

65. AKRAM MF, ALI, M. R., KADIROVA Z (2023b) Role of consumer digital literacy in consumer behavioral intentions: Mediating effect of perceived ease of use, usefulness and customer engagement. Multidisciplinary J Educational Res 13:104–130
66. ALI MR, AKRAM MF, TASHTAYEVA URKINBAYEVTA, S. K., ARIPDJANOV GT (2023) The Pathways to Achieve Green Economy and The Process of Sustainable Development Goals, vol 2. SGS-Engineering & Sciences
67. Ali MR, IN GREEN HUMAN RESOURCE MANAGEMENT: UNDERSTANDING THE MEDIATING ROLE OF EMPLOYEE ENGAGEMENT IN ACHIEVING ENVIRONMENTAL SUSTAINABILITY (2025) Dr. Liga Peiseniece, Dr. Elina Mikelsone ARTIFICIAL INTELLIGENCE. Zenodo. Available at: https://doi.org/10.5281/zenodo.15193336
68. Ali MR, IN GREEN HUMAN RESOURCE MANAGEMENT: UNDERSTANDING THE MEDIATING ROLE OF EMPLOYEE ENGAGEMENT IN ACHIEVING ENVIRONMENTAL SUSTAINABILITY (2025) Dr. Liga Peiseniece, Dr. Elina Mikelsone ARTIFICIAL INTELLIGENCE. Zenodo. Available at: https://doi.org/10.5281/zenodo.15193336
69. Alimova Zebo1 Muhammad Ramzan Ali(PhD)2, Bakhronkulova Laylo3, Farzona Mirsodikova4, 5Irkinov Rasuljon (2025) THE ROLE OF AI IN DEVELOPING GLOBAL CYBERSECURITY POLICIES. Zenodo. Available at: https://doi.org/10.5281/zenodo.14899972
70. Ali MR, Dr Muhammad Bilal Qureshi, Dr Adnan Manzoor (2025) CYBER SECURITY AND ARTIFICIAL INTELLIGENCE: REVOLUTIONARY TRENDS CREATING THE FUTURE. Zenodo. Available at: https://doi.org/10.5281/zenodo.14911281
71. Kadirovich XC, Shah ASA, Ali MR, Khasanovna VM (2026) The Moderating Role of Staff Training and Mediating Role of Venue Reputation in Accessible MICE Tourism: Effects on. https://doi.org/10.21203/rs.3.rs-11106800/v1. Business Event Satisfaction
72. Ali MR (n.d.) Artificial Intelligence in Green Human Resource Management: Examining the Mediating Role of Employee Engagement in Attaining Organizational Sustainability. Research Square. Available at: https://doi.org/10.21203/rs.3.rs-11412497/v1
73. Lyster R, Ranta L (1997) Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. Stud Second Lang Acquisition 19(1):37–66. https://doi.org/10.1017/S0272263197001034
74. *Meritan C (2022) The longitudinal impact of self-reflection and integrated pronunciation instruction on L2 French learners' production. J Second Lang Pronunciation 8(1):116–143. https://doi.org/10.1075/jslp.21022.mer
75. *Muranoi H (2000) Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. Lang Learn 50(4):617–673. https://doi.org/10.1111/0023-8333.00142
76. *Nguyen HTM, Nguyen HLT, Vo NM, Huynh NK (2025) The combined effects of task repetition and post-task teacher-corrected transcribing on complexity, accuracy and fluency of L2 oral performance. IRAL - Int Rev Appl Linguistics Lang Teach 63(2):1317–1345. https://doi.org/10.1515/iral-2023-0128
77. Nicol DJ, Macfarlane-Dick D (2006) Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Stud High Educ 31(2):199–218. https://doi.org/10.1080/03075070600572090
78. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L et al (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ 372:n71. https://doi.org/10.1136/bmj.n71
79. Panadero E, Jonsson A, Botella J (2017) Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. Educational Res Rev 22:74–98. https://doi.org/10.1016/j.edurev.2017.08.004

80. Philp J, Duchesne S (2016) Exploring engagement in tasks in the language classroom. Annu Rev Appl Linguist 36:50–72. https://doi.org/10.1017/S0267190515000094
81. *Piurko E (2026) Creative communication in English for specific purposes through multifaceted speaking in panel discussions. Creativity Stud 19(1):202–210. https://doi.org/10.3846/cs.2026.24678
82. Plonsky L, Oswald FL (2014) How big is big? Interpreting effect sizes in L2 research. Lang Learn 64(4):878–912. https://doi.org/10.1111/lang.12079
83. Priem J, Piwowar H, Orr R (2022) OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv preprint. 2205.01833 https://doi.org/10.48550/arXiv.2205.01833
84. *Qureshi B, Roseberry J, Qureshi Z (2020) Boosting intersubjectivity by digital video recording project in EFL classes. Scope: Journal of English Language Teaching 4(1): 20–27. https://doi.org/10.30998/scope.v4i01.4977
85. *Rif SC, Tan KH (2021) Debriefing strategy in online primary ESL classroom during COVID-19 pandemic: A case study. Int J Adv Res Educ Soc 3(4). https://doi.org/10.55057/ijares.2021.3.4.8
86. *Rif SC, Tan KH, Khairuddin KF (2023) Debriefing in online primary ESL classrooms during the COVID-19 pandemic: A case study. J Lang Teach Res 14(1):224–230. https://doi.org/10.17507/jltr.1401.23
87. Robinson P (2001) Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. Appl Linguist 22(1):27–57. https://doi.org/10.1093/applin/22.1.27
88. Rudolph JW, Simon R, Dufresne RL, Raemer DB (2006) There's no such thing as nonjudgmental debriefing: A theory and method for debriefing with good judgment. Simul Healthc 1(1):49–55. https://doi.org/10.1097/01266021-200600110-00006
89. *Saeedi M (2023) Task implementation and EFL learners' attention to form: Examining the effects of pre-task planning and post-task transcription. Issues Lang Teach 12(1):67–96. https://doi.org/10.22054/ilt.2023.67865.697
90. Sawyer T, Eppich W, Brett-Fleegler M, Grant V, Cheng A (2016) More than one way to debrief: A critical review of healthcare simulation debriefing methods. Simul Healthc 11(3):209–217. https://doi.org/10.1097/SIH.0000000000000148
91. Schmidt RW (1990) The role of consciousness in second language learning. Appl Linguist 11(2):129–158. https://doi.org/10.1093/applin/11.2.129
92. *Shih R-C (2010) Blended learning using video-based blogs: Public speaking for English as a second language students. Australasian J Educational Technol 26(6):883–897. https://doi.org/10.14742/ajet.1048
93. Skehan P (1996) A framework for the implementation of task-based instruction. Appl Linguist 17(1):38–62. https://doi.org/10.1093/applin/17.1.38
94. Skehan P (1998) A cognitive approach to language learning. Oxford University Press, Oxford
95. Skehan P, Foster P (1997) Task type and task processing conditions as influences on foreign language performance. Lang Teach Res 1(3):185–211. https://doi.org/10.1177/136216889700100302
96. *Skehan P, Bei X, Li Q, Wang Z (2012) The task is not enough: Processing approaches to task-based performance. Lang Teach Res 16(2):170–187. https://doi.org/10.1177/1362168811428414
97. Small H (1973) Co-citation in the scientific literature: A new measure of the relationship between two documents. J Am Soc Inform Sci 24(4):265–269. https://doi.org/10.1002/asi.4630240406
98. *Song B (2017) Effects of task repetition and self-reflection on EFL learners' attentional allocation and speaking performance. Engl Teach 72(4):81–103. https://doi.org/10.15858/engtea.72.4.201712.81

99. *Sulistyo T, Lutviana R, Purnawati M, Widiastuti O, Ginting D (2026) From production to reflection: Integrating video project-based learning and dialogic pedagogy in tertiary EFL speaking. Engl Lang Teach Educational J 9(2). https://doi.org/10.12928/eltej.v9i2.16237
100. Šulovská D (2026) Role of debriefing in enhancing ESP classroom scenario-based activities. CASALC Rev 15(2):4–19. https://doi.org/10.5817/CASALC2025-2-1
101. *Tsai C-C, Piamsai C (2025) The use of peer assessment and non-scripted role-play activities in improving Thai EFL high school students' English oral performance. rEFLections 32(1):126–153. https://doi.org/10.61508/refl.v32i1.279180
102. *van Compernolle RA, Ballesteros Soria N (2020) Developing interactional repertoires in the classroom through dynamic strategic interaction scenarios. Vigo Int J Appl Linguistics 17. https://doi.org/10.35869/vial.v0i17.1468
103. *Velásquez-Hoyos AP (2021) Theme-based teaching to promote oral fluency in a university in Colombia. HOW 28(2):97–119. https://doi.org/10.19183/how.28.2.555
104. *Villamizar AG, Mejía G (2019) Fostering learner autonomy and critical reflection through digital video-journals in a university foreign language course. Reflective Pract 20(2):187–200. https://doi.org/10.1080/14623943.2019.1575195
105. *Vo NLG, Le DH (2026) Task-based language teaching and e-portfolios in enhancing EFL students' autonomous learning competencies and speaking performance. VNUHCM J Social Sci Humanit 10(3). https://doi.org/10.32508/vnuhcmjssh.v10i3.1308
106. Vygotsky LS (1978) Mind in society: The development of higher psychological processes. Harvard University Press, Cambridge, MA
107. Wald HS, Borkan JM, Taylor JS, Anthony D, Reis SP (2012) Fostering and evaluating reflective capacity in medical education: Developing the REFLECT rubric for assessing reflective writing. Acad Med 87(1):41–50. https://doi.org/10.1097/ACM.0b013e31823b55fa
108. Waltman L, van Eck NJ, Noyons ECM (2010) A unified approach to mapping and clustering of bibliometric networks. J Informetrics 4(4):629–635. https://doi.org/10.1016/j.joi.2010.07.002
109. *Wang P (2026) Marrying task-based language teaching and translanguaging: A curriculum-level case study of a Chinese Tier-3 vocational university. Curriculum J Adv online publication. https://doi.org/10.1002/curj.70091
110. Willis J (1996) A framework for task-based learning. Longman, Harlow
111. *Winke P (2014) Formative, task-based oral assessments in an advanced Chinese-language class. In Technology-mediated TBLT: Researching technology and tasks, edited by M. González-Lloret and L. Ortega, 263–294. Amsterdam: John Benjamins. https://doi.org/10.1075/tblt.6.10win
112. *Yin J (2017) The effects of multi-source feedback on EFL students' reflection of English speaking skills. J Korea Engl Educ Soc 16(3):155–173. https://doi.org/10.18649/jkees.2017.16.3.155
113. *Yu H, Li Z, Köhler ES, van Merriënboer JJG, Asoodar M (2025) How do medical students make sense of internal and external feedback to enhance their Dutch communication skills? BMC Med Educ 25:256. https://doi.org/10.1186/s12909-025-06845-0
114. Zimmerman BJ (2002) Becoming a self-regulated learner: An overview. Theory Into Pract 41(2):64–70. https://doi.org/10.1207/s15430421tip4102_2
115. Zupic I, Čater T (2015) Bibliometric methods in management and organization. Organizational Res Methods 18(3):429–472. https://doi.org/10.1177/1094428114562629

Figures

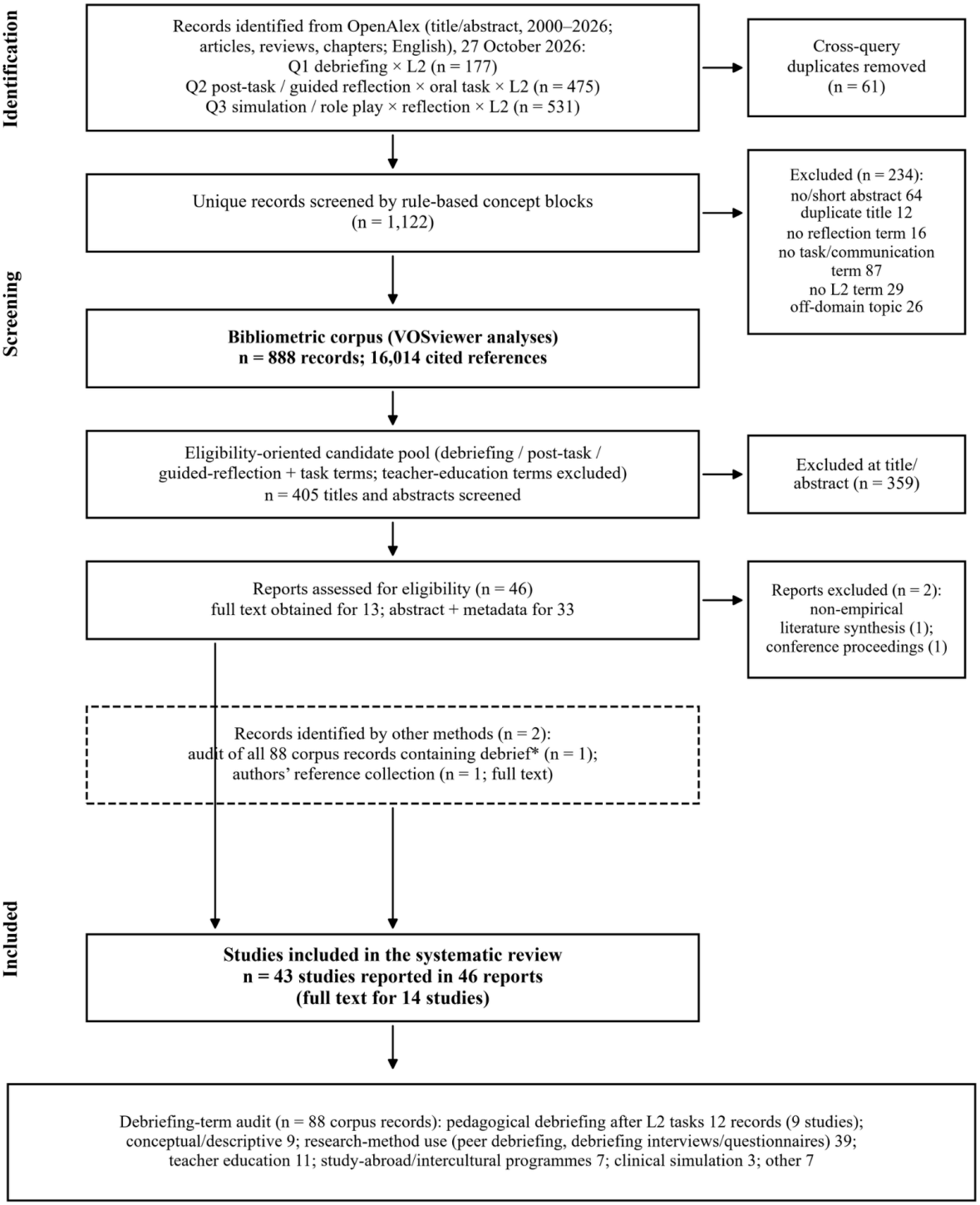


Figure 1

PRISMA 2020 flow diagram of identification, screening, and inclusion, with the debriefing-term audit

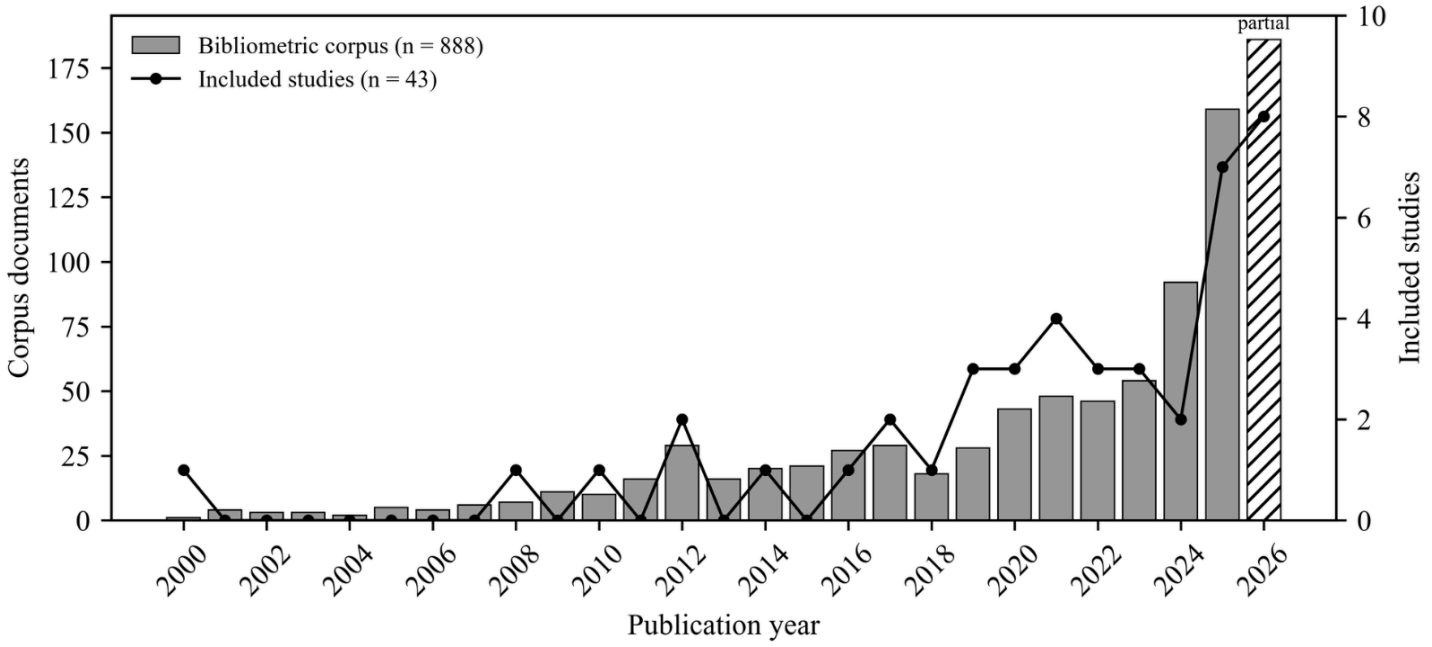


Figure 2

Annual output of the bibliometric corpus (bars; n = 888) and of the studies included in the systematic review (line; n = 43); 2026 covers 1 January–27 October

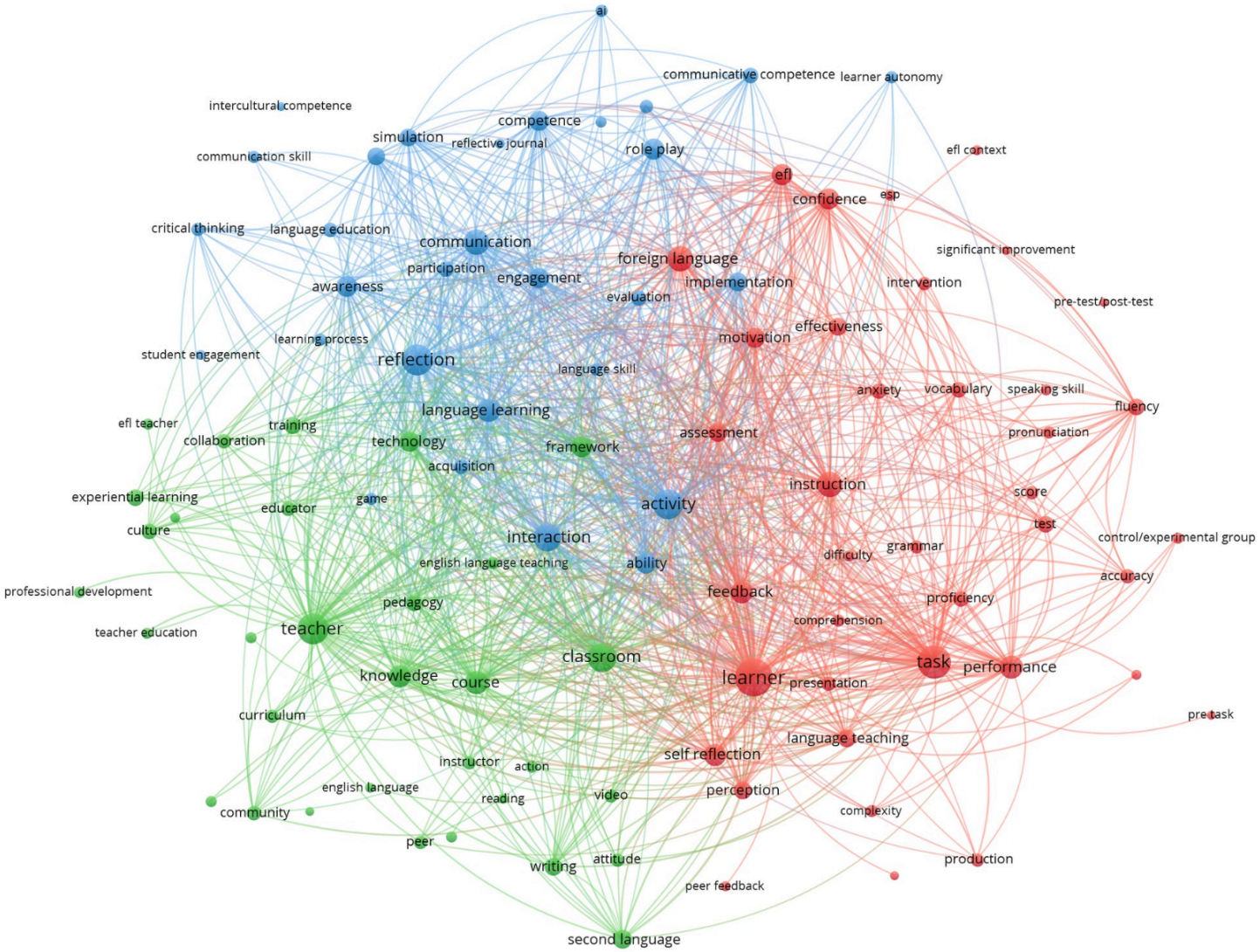


Figure 3

Term co-occurrence network (VOSviewer 1.6.21; 101 terms with at least 25 occurrences; binary counting; node size = occurrences; colour = cluster)

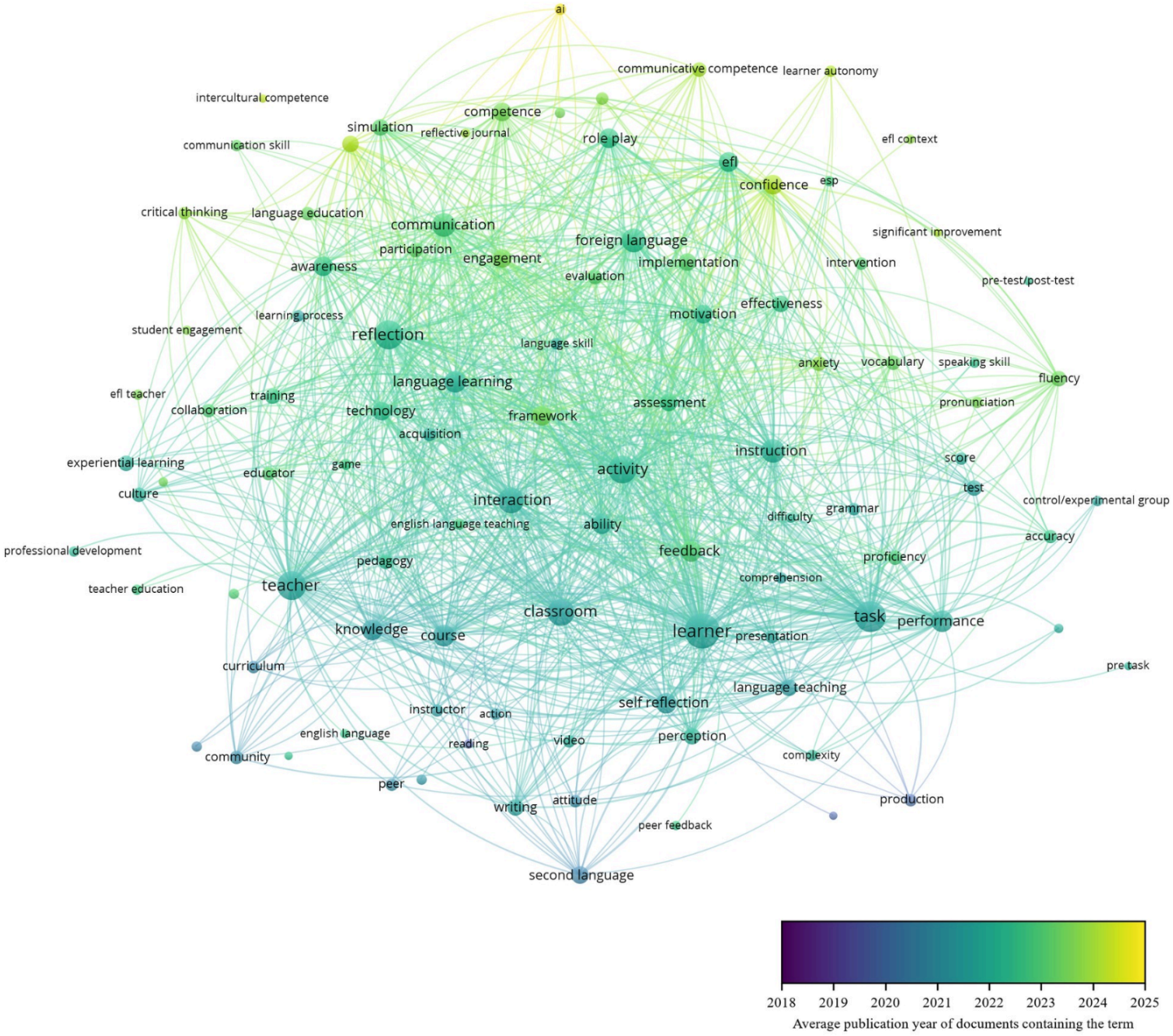


Figure 4

Overlay visualisation of the term map coloured by average publication year (VOSviewer viridis scale fixed at 2018–2025)

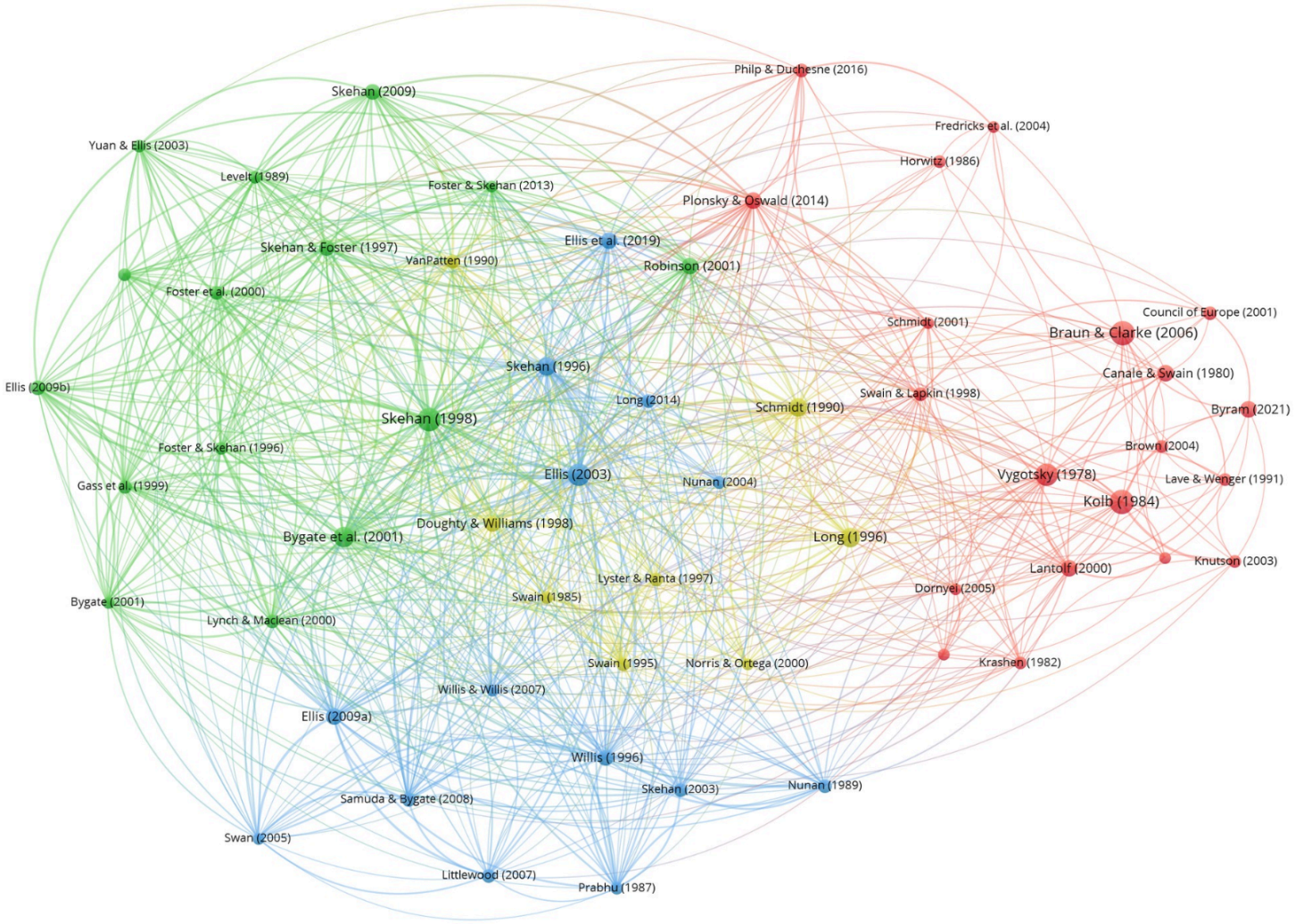


Figure 5

Co-citation network of cited references (VOSviewer 1.6.21; 57 works cited by at least eight corpus documents; link strength = number of co-citing documents)

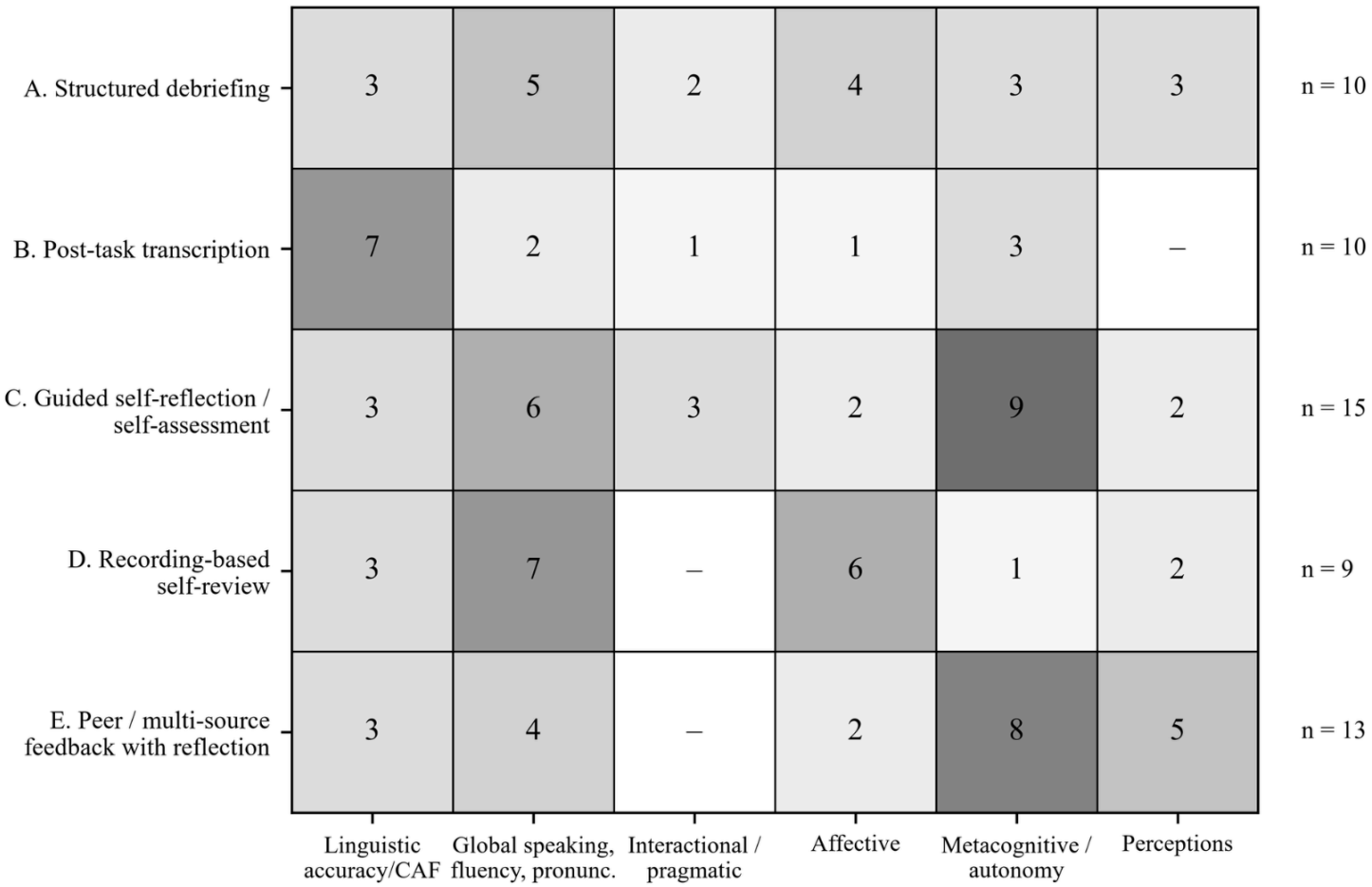


Figure 6

Evidence map: number of included studies by reflective modality and outcome domain (studies may contribute to several cells)

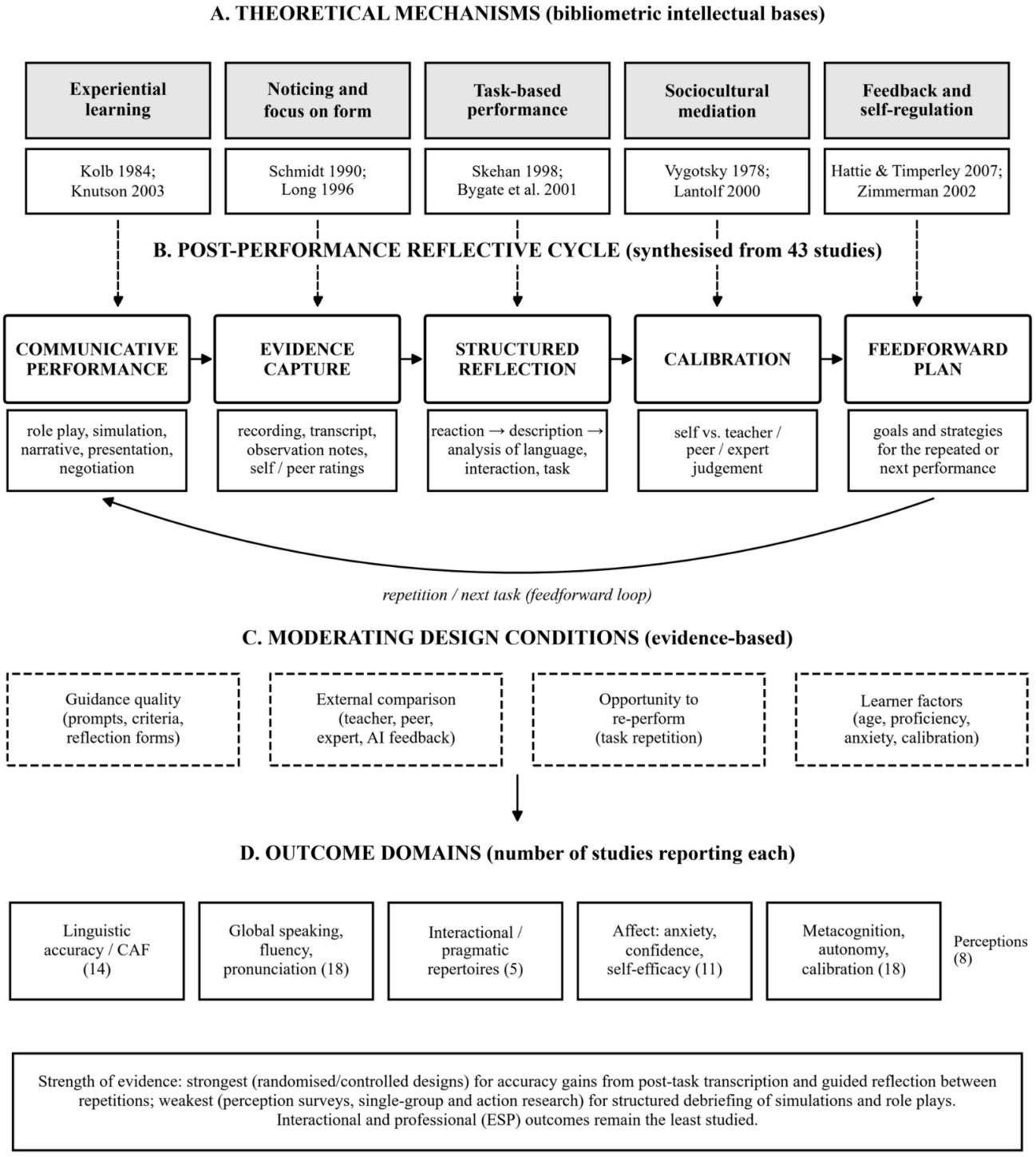


Figure 7

Post-performance reflective cycle for L2 and ESP communicative tasks: theoretical mechanisms, phases, moderating conditions, and outcome domains