



# Generative AI and the Future of Needs Analysis in LSP Course Design: A Systematic Review

Wiam Hadji

Laboratory of Pragmatics and Discourse Analysis, University of El Oued, Algeria

## Abstract

*Needs analysis is the methodological cornerstone of Language for Specific Purposes (LSP) course design, yet it remains slow, costly and, in many institutional settings, methodologically thin. Generative artificial intelligence now offers a capacity to analyse occupational discourse, simulate stakeholder input and draft syllabi at a speed that conventional procedures cannot match. This systematic review, conducted and reported in accordance with the PRISMA 2020 statement, examines studies published between January 2020 and September 2026 that investigate the use of large language models at the needs analysis and course design stages of LSP and English for Specific Purposes provision. Searches of Scopus, ScienceDirect, ERIC and Google Scholar identified 968 records; after removal of 205 duplicates and screening of the remaining 763 against predefined criteria, 39 studies were included. The corpus is recent and unevenly distributed: no eligible study predates 2024, course design and materials design account for eleven studies each, while only five address needs analysis. Three applications emerge: automated analysis of occupational and academic discourse, AI-simulated stakeholder consultation, and the drafting of syllabi and materials. Validation is concentrated where outputs can be checked against an existing text and is largely absent where they cannot, which makes needs analysis the weakest-evidenced stage. The paper proposes a four-phase hybrid model that positions generative AI as a preparatory instrument within established LSP methodology rather than as a replacement for it.*

**Keywords:** generative AI; needs analysis; Language for Specific Purposes; large language models; course design; systematic review

## 1. Introduction

Language for Specific Purposes (LSP) is defined less by its content than by its procedure. What distinguishes an LSP course is that its objectives are derived empirically from an analysis of what learners will have to do with the language, and needs analysis is the instrument of that derivation [1], [2]. Successive frameworks have extended target and present situation analysis with task-based needs analysis and means analysis [3], [4], but one requirement has remained constant: needs statements must be triangulated across sources and methods and validated by domain insiders before they are converted into syllabus specifications. In practice this requirement is frequently unmet. An audit of needs analyses for specialised populations found widespread reliance on a single source and a single method [5], and a decade-scale review of empirical ESP research confirms that methodological innovation has concentrated on delivery rather than on the procedures that precede it [6].

Generative artificial intelligence has entered this gap: large language models summarise occupational discourse and draft syllabi in minutes. This review establishes what is known about that use and converts the evidence into a defensible procedure. It asks three questions.

RQ1. In what ways has generative AI been applied to needs analysis and course design in LSP and ESP contexts in studies published between 2020 and 2026?

RQ2. What evidence do these studies report concerning the accuracy of AI-generated outputs, and what human validation procedures, if any, do they document?

RQ3. What concerns recur across the corpus, and what model of practice do the findings collectively support?

### 1.1 Previous Relevant Reviews and the Gap Addressed

Four reviews of generative AI in language education have appeared since 2024 [7], [8], [9], [10], together with two of the wider educational literature [11], [12] and one decade-scale review of ESP research [6]. Three limitations are common to them. Each takes the learner rather than the course as the unit of



analysis; each reports what generative AI is used for rather than whether its outputs were validated; and none is specific to LSP, where the requirement for validated needs evidence is most exacting.

## 2. Conceptual Background

### 2.1 What Needs Analysis Requires

Needs analysis is an evidential procedure rather than a survey, and it imposes three requirements. Data must originate in the target domain, as workplace texts, observation or insider testimony; findings must be triangulated across sources and methods; and needs statements must be validated by people who perform the target tasks, because the language practitioner is by definition an outsider to the domain being analysed [3], [5]. A procedure that meets the first two requirements but not the third produces a plausible syllabus rather than a warranted one, and plausibility is precisely what generative models are optimised to produce.

### 2.2 What Large Language Models Demonstrably Do

Evidence about model capability in specialised discourse is mixed. In a systematic comparison with corpus methods, ChatGPT performed acceptably when assigning keywords to semantic categories but poorly on concordance analysis, generating false inferences and in some instances altering the input data [13]. The question for LSP is therefore not whether generative AI is accurate enough to replace needs analysis, but where in the procedure it can contribute and what must be checked when it does.

## 3. Methodology

### 3.1 Design, Reporting Standard and Protocol

The review follows the PRISMA 2020 statement [14] and synthesises a heterogeneous, predominantly qualitative corpus narratively rather than statistically.

The protocol, specifying the research questions, eligibility criteria, search strings, coding framework and synthesis procedure, was written before searching began and was not amended. It was not registered in PROSPERO, whose scope is restricted to reviews with health-related outcomes, and no equivalent register exists for applied linguistics; the protocol is available from the author.

### 3.2 Eligibility Criteria

Eligibility was defined using the population, concept and context structure recommended for reviews that do not compare interventions [15], extended with publication type and indexing (Table 1). The concept dimension admits both studies that apply a model to a design task and studies that investigate how practitioners use it at that stage. The context dimension carries a secondary clause admitting studies from adjacent fields where they supply design-stage evidence the language-education literature lacks: the elicitation of needs from simulated stakeholders, and the measurement of generated output against an external reference.

**Table 1.** Inclusion and exclusion criteria

| Dimension | Included                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Excluded                                                                                                                             |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|
| Concept   | Empirical studies of the use of generative AI or large language models at any stage of the needs analysis or course design cycle (target or present situation analysis, task inventories, genre and discourse analysis for course purposes, syllabus or materials specification, assessment specification). This covers both studies that apply a model to a design task and studies that investigate how practitioners use it, or what they report about using it, at those stages | Learner-facing applications only (writing feedback, conversational practice, assessment of learners) with no course-design component |



| Dimension           | Included                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Excluded                                                                                                                                               |
|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| Context             | Primary: LSP, ESP, EAP, vocational or professional language provision, any language, any educational sector. Secondary: studies outside language education were eligible only where they supply design-stage evidence that the language-education literature does not yet provide, namely the elicitation of needs from simulated rather than real stakeholders, or the measurement of generated output against an external reference such as an existing curriculum | General language education with no specific-purposes orientation; studies outside language education addressing procedures already evidenced within it |
| Publication type    | Peer-reviewed journal articles and peer-reviewed conference papers reporting original empirical data or a documented analytical procedure. Opinion pieces, forum articles and commentaries were excluded at full-text stage even where the journal indexes them as research                                                                                                                                                                                          | Editorials, opinion pieces, unrefereed preprints, theses, secondary reviews (retained for reference-list screening only)                               |
| Indexing            | Published in a journal indexed in Scopus or the Web of Science Core Collection, verified against the Scopus Source List and the Clarivate Master Journal List                                                                                                                                                                                                                                                                                                        | Publications in non-indexed outlets, repositories and preprint servers                                                                                 |
| Period and language | Published January 2020 – 3 September 2026 (date of final search); reported in English                                                                                                                                                                                                                                                                                                                                                                                | Published outside the window; not available in English                                                                                                 |

### 3.3 Information Sources and Search Strategy

Scopus, ScienceDirect, ERIC and Google Scholar were searched on 1 and 3 September 2026, the last restricted to the first 100 relevance-ranked records. Because the terminology of the field is not standardised, a recall-first strategy combined a technology block (generative AI, generative artificial intelligence, large language model(s), LLM, ChatGPT, GPT-4, Gemini, Copilot) with a domain block (English for Specific Purposes, specific purposes, ESP, LSP, EAP, professional communication, workplace language, vocational English) across sixteen Scopus strings; a precision check added a process block (needs analysis, needs assessment, target situation analysis, course, curriculum and syllabus design, materials development, task analysis). The searches were supplemented by forward and backward citation chaining and by hand-searching ten specialist journals. In Scopus and ScienceDirect the publication-year filter is populated from the records a search returns, and for every string the earliest year offered was 2023: the searches covered the full window but returned nothing earlier, a result confirmed independently by ERIC and the hand-search. Indexing status was verified for every included venue against the Scopus Source List and the Clarivate Master Journal List.

### 3.4 Selection Process

Records were de-duplicated by digital object identifier and normalised title, then screened by the author in two complete passes separated by an interval. Discrepancies between passes were resolved by returning to the written criteria rather than to the earlier decision, so that the criteria, not the first judgement, remained the reference point.

### 3.5 Coding Framework, Quality Appraisal and Synthesis

Included studies were charted against a seven-dimensional framework: context, design stage, research method, theoretical grounding, model and input, validation procedure and reported outcomes. The last three capture what is distinctive about generative AI as a research instrument: what the model was given, what it was asked to produce, and whether anyone checked the result. Methodological quality was appraised with the Mixed Methods Appraisal Tool [16], supplemented by four transparency



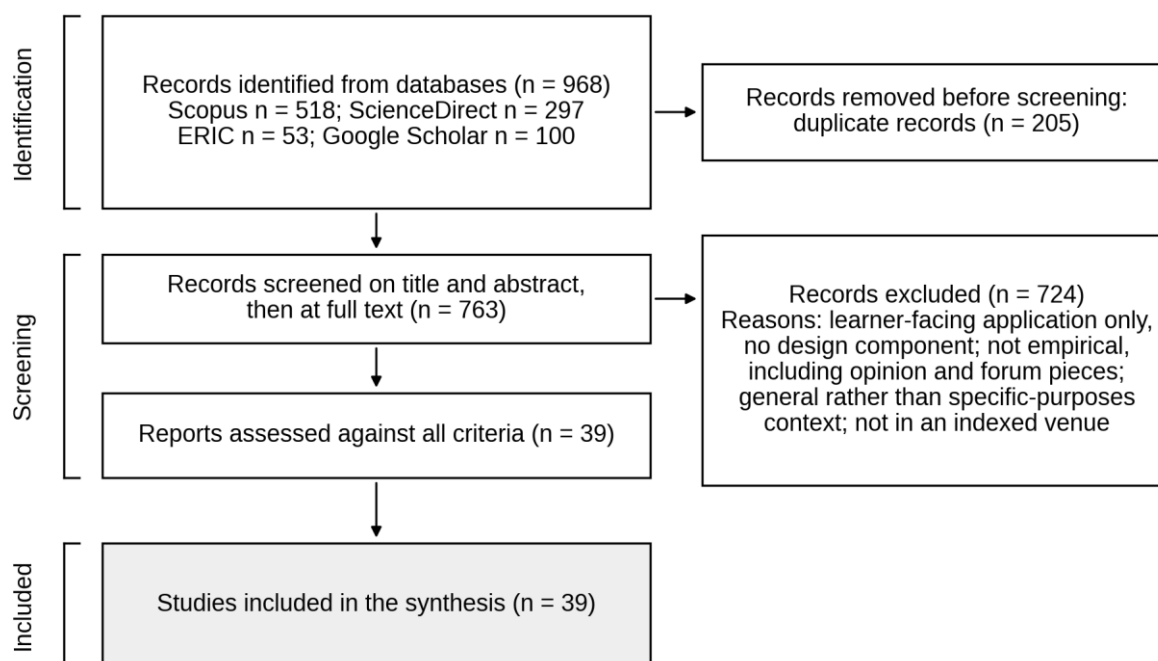
indicators: model and version reported, prompts reported verbatim, repeated runs documented, and outputs validated against a benchmark.

Studies were synthesised thematically through line-by-line coding, descriptive themes and analytical themes [17]. Descriptive themes were grouped by the function the model was asked to perform rather than by the technology used; analytical themes, asking what the findings imply for a defensible needs analysis procedure, constitute the model in Section 5. Because the included studies report overwhelmingly positive outcomes, the possible under-representation of negative or null findings is treated as a limitation of the evidence.

## 4. Results

### 4.1 Study Selection

The searches returned 968 records: 518 from Scopus, 297 from ScienceDirect, 53 from ERIC and 100 from Google Scholar. After removal of 205 duplicates (158 within Scopus, 28 within ScienceDirect and 19 across databases), 763 records were screened and 724 excluded, the large majority because the application was learner-facing, as in writing feedback, conversational practice or automated scoring, with no design component. Thirty-nine studies met all criteria and were included (Fig. 1). Strings naming the design stage returned markedly fewer records than strings naming the field: design-stage work exists but is not indexed as such.



**Fig. 1.** PRISMA 2020 flow diagram of the study selection process

### 4.2 Distribution of the Corpus

Table 2 reports the distribution of the 39 studies. No eligible study was published before 2024: three appeared in 2024, eight in 2025 and twenty-eight in 2026, so that design-stage adoption followed the conversational interface released in late 2022, and the review describes a literature roughly two years old. Course or curriculum design and materials design dominate, while needs analysis, the stage from which every other decision is supposed to derive, accounts for five studies, three of them in a language-education context. Thirty-five studies concern English, and a single study, on Arabic for academic and professional purposes [18], constitutes the entire indexed evidence base for LSP beyond English. By design stage, course or curriculum design is addressed by [19]–[29]; materials and task design by [30]–[40]; genre and discourse analysis for course purposes by [41]–[47]; assessment specification by [48]–[52]; and needs analysis by [18], [53]–[56].



**Table 2.** Distribution of the included studies (n = 39)

| Dimension              | Category                                                               | n  |
|------------------------|------------------------------------------------------------------------|----|
| Year of publication    | 2020–2023                                                              | 0  |
|                        | 2024                                                                   | 3  |
|                        | 2025                                                                   | 8  |
|                        | 2026                                                                   | 28 |
| Design stage addressed | Materials and task design                                              | 11 |
|                        | Course or curriculum design                                            | 11 |
|                        | Genre and discourse analysis for course purposes                       | 7  |
|                        | Assessment specification                                               | 5  |
|                        | Needs analysis                                                         | 5  |
| Context                | English for Specific Purposes (occupational, vocational, professional) | 17 |
|                        | English for Academic Purposes                                          | 18 |
|                        | Adjacent fields, admitted under the secondary context criterion        | 3  |
|                        | Languages for specific purposes other than English                     | 1  |

#### **4.3 Characteristics, Quality and Certainty of the Evidence**

The corpus is small and recent. Perception studies, in which practitioners or teachers report what they believe generative AI can do, outnumber studies that evaluate what it actually produces. Only a minority supplied the model with authentic domain texts; the remainder relied on the model's parametric knowledge of a profession, a materially different and considerably weaker evidential basis.

On the Mixed Methods Appraisal Tool the corpus is adequate, but on the transparency indicators it is weak: prompts are more often described than reproduced, and repeated runs are rarely documented. A review restricted to SSCI-indexed studies reports the same deficit across language education [10], and guidance in applied linguistics journals remains uneven [57], which points to a reporting apparatus designed for stable instruments rather than to individual carelessness. Certainty, characterised qualitatively because outcome measures are not comparable, is moderate for discourse analysis, low for syllabus drafting and very low for simulated stakeholders.

#### **4.4 RQ1, Theme 1: Automated Analysis of Occupational and Academic Discourse**

The best-evidenced application uses a model to perform target situation analysis on texts rather than on people. ChatGPT-4 was used to annotate rhetorical moves in CEO statements from corporate social responsibility reports: after prompt design and validation on 30 statements, the optimised prompt achieved 87.14% accuracy on a further 50, but double annotation revealed an intra-model inconsistency rate of 32.87% [41]. In move-step analysis of 100 applied linguistics research article introductions, few-shot learning and prompt refinement produced statistically significant but small improvements, while fine-tuning raised accuracy to 92.3% for moves and 80.2% for steps [42]. Accuracy is therefore a property of the prompt and the fine-tuning as much as of the model, and an unreported prompt is an unreported method. Building a small corpus retains its value because it supplies evidence the model cannot [43], and AI-assisted text is itself changing the professional genres a corpus would record [45].

#### **4.5 RQ1, Theme 2: AI-simulated Stakeholder Consultation**

Because access to professionals constrains most needs analyses, prompting a model to respond as a nurse or an engineer is the most attractive application and the least supported. ESP practitioners report confidence in the model's account of target-situation requirements [53], yet no study in a language-



education context has compared simulated with real stakeholders. The direct tests come from adjacent fields, where large language models have served as simulated interviewees in engineering design [54] and been compared on automated customer needs analysis [55]. Synthetic respondents reproduce aggregate distributions [58] but are unstable by subgroup [59], so a simulated stakeholder gives the modal account of an occupation and misses local practice and tacit knowledge.

#### 4.6 RQ1, Theme 3: Generation of Draft Syllabi and Task Banks

Twenty-two studies address materials, course or curriculum design. Practitioners describe prompting strategies and self-imposed ethical limits [31]; curricula have been redesigned with ChatGPT [48], evaluated with pre- and post-tests [20], [25], and compared against existing university programmes [21]. A three-year case of a non-specialist designing a bio-health English course credits AI with the domain confidence it gave an outsider to the discipline [19], and an ethnography of an ESP instructor documents the labour of converting tacit pedagogical knowledge into instructions a model can act on [30]. Both describe one risk: plausible domain content reaches the person least equipped to detect its errors.

#### 4.7 RQ2: Accuracy Evidence and Validation Procedures

Validation clusters where output can be checked against something that already exists. It takes four forms: benchmarking against manual analysis [41], [42]; comparison of machine scores against human raters, as in a GPT-based EAP speaking assessment analysed with many-facet Rasch measurement [52]; implementation with learners [23], [25], [29]; and comparison with existing curricula [21]. It is absent where no prior text exists: a generated needs statement can be checked only against a human stakeholder, and no study in language education has done so.

#### 4.8 RQ3: Recurring Concerns Across the Corpus

Four concerns recur: data currency, since a model's knowledge of an occupation lags professional change; disciplinary accuracy, hardest to verify for the non-specialist who most relies on it; reproducibility, undermined by non-deterministic outputs and unreported prompts; and the confidentiality of workplace documents supplied to a commercial model.

### 5. Discussion: A Hybrid Needs Analysis Model

The findings support specification rather than prohibition or uncritical adoption, in four phases (Fig. 2). Phase 0 establishes governance and a log of model, version and date. Phase 1 uses the model for preparatory analysis of authentic domain texts, each prompt run more than once; its outputs are hypotheses only. Phase 2 tests every provisional item against domain insiders and the corpus and adds what the model omitted. Phase 3 writes the syllabus with provenance tagged per item.

|                                             |                                                                                                                                                                 |                          |
|---------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| Phase 0: Scoping and governance             | Define the target situation and stakeholders; obtain ethical clearance; decide what may lawfully be submitted to a model; open a log of model, version and date | Procedural record        |
| Phase 1: AI-assisted preparatory analysis   | Upload authentic domain texts; generate a provisional genre map, task inventory and draft instruments; run each prompt more than once                           | Hypotheses only          |
| Phase 2: Human validation and triangulation | Test every provisional item against domain insiders and the corpus; confirm, amend or discard; add what the model did not produce                               | Evidence with its source |
| Phase 3: Specification and audit            | Write the syllabus with provenance tagged per item; archive prompts, model version and validation records; set a re-validation date                             | Warranted syllabus       |

Validation never reverses direction: a model is never used to adjudicate between human sources.

**Fig. 2.** The four-phase hybrid needs analysis model



Provenance is tracked per item because a course that is 80% validated is not 80% defensible if no one can say which 20% is not. Where existing reviews map application types [7], [8], [10], this review reports the distribution of validation, and finds it inverted with respect to need. The procedure redistributes effort rather than reducing it: time saved in Phase 1 funds the stakeholder contact first sacrificed under institutional deadlines, making triangulated needs analysis affordable in resource-constrained provision. The model assumes that domain insiders are reachable, that authentic texts may lawfully be uploaded, and that the institution will accept an audit trail. Where one fails, the procedure degrades into something weaker but declarable rather than failing silently.

## 6. Limitations

Five limitations apply. Screening and coding were conducted by one reviewer in two passes; the protocol was not prospectively registered; the corpus is small and weighted towards perception studies; only English-language publications were included; and model development outpaces the publication cycle.

## 7. Conclusion and Directions for Research

Generative AI is already used at the design stage of LSP, but the evidence for its accuracy is concentrated where it is least needed and absent where it is most needed. Three priorities follow: validation studies comparing simulated with real stakeholders; outcome studies comparing hybrid and conventional needs analysis against learner performance; and a minimum reporting standard covering model, version, date, prompts, number of runs and validation procedure. Until then, the defensible position is to use the tool at the preparatory stage, validate every item it produces, and document how that validation was carried out.

## Declarations

**Ethics statement.** This study analysed published literature only and involved no human participants; ethical approval was therefore not required.

**Declaration of conflicting interests.** The author declares no potential conflicts of interest with respect to the research, authorship or publication of this article.

**Funding.** This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

**Data availability.** The following are available from the author on reasonable request: the charting table recording the context, design stage, method, model, input and validation procedure of each included study; the comparative table of previous reviews; the search strings as executed in each database, with hit counts and dates; and the screening record.

**Declaration of generative AI use.** During the preparation of this manuscript the author used a generative AI assistant to support literature identification (designing the Boolean strings), drafting and language editing. All sources were independently verified by the author, all inclusion decisions were reviewed and confirmed by the author against the criteria in Table 1, and the author reviewed and edited the content and takes full responsibility for the published paper.

## REFERENCES

References marked with an asterisk are the 39 studies included in the review.

- [1] Hutchinson T., Waters A., "English for specific purposes: A learning-centred approach", Cambridge, Cambridge University Press, 1987.
- [2] Dudley-Evans T., St John M. J., "Developments in English for specific purposes: A multi-disciplinary approach", Cambridge, Cambridge University Press, 1998.
- [3] Long M. H. (Ed.), "Second language needs analysis", Cambridge, Cambridge University Press, 2005.
- [4] Basturkmen H., "Developing courses in English for specific purposes", Basingstoke, Palgrave Macmillan, 2010.
- [5] Serafini E. J., Lake J. B., Long M. H., "Needs analysis for specialized learner populations: Essential methodological improvements", *English for Specific Purposes*, 40, pp. 11–26, 2015.
- [6] Yan H., "Trends in empirical research in English for Specific Purposes: A systematic review of SSCI-indexed journal articles (2014–2023)", *SAGE Open*, 15(1), 2025.



- [7] Law L., "Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review", *Computers and Education Open*, 6, Article 100174, 2024.
- [8] Lee S., Choe H., Zou D., Jeon J., "Generative AI (GenAI) in the language classroom: A systematic review", *Interactive Learning Environments*, 33(1), pp. 335–359, 2025.
- [9] Li B., Tan Y. L., Lowell V., "Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching", *Computers and Education: Artificial Intelligence*, 2025.
- [10] Wang Y., Zhang T., Yao L., Seedhouse P., "A scoping review of empirical studies on generative artificial intelligence in language education", *Innovation in Language Learning and Teaching*, pp. 1–28, 2025.
- [11] Shi Y., Yu K., Dong Y., Chen F., "Large language models in education: A systematic review of empirical applications, benefits, and challenges", *Computers and Education: Artificial Intelligence*, 10, Article 100529, 2025.
- [12] Yan L., Sha L., Zhao L., Li Y., Martinez-Maldonado R., Chen G., Li X., Jin Y., Gašević D., "Practical and ethical challenges of large language models in education: A systematic scoping review", *British Journal of Educational Technology*, 55(1), pp. 90–112, 2024.
- [13] Curry N., Baker P., Brookes G., "Generative AI for corpus approaches to discourse studies: A critical evaluation of ChatGPT", *Applied Corpus Linguistics*, 4(1), Article 100082, 2024.
- [14] Page M. J., McKenzie J. E., Bossuyt P. M., Boutron I., Hoffmann T. C., Mulrow C. D., Shamseer L., Tetzlaff J. M., Akl E. A., Brennan S. E., Chou R., Glanville J., Grimshaw J. M., Hróbjartsson A., Lalu M. M., Li T., Loder E. W., Mayo-Wilson E., McDonald S., ... Moher D., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews", *BMJ*, 372, Article n71, 2021.
- [15] Peters M. D. J., Marnie C., Tricco A. C., Pollock D., Munn Z., Alexander L., McInerney P., Godfrey C. M., Khalil H., "Updating methodological guidance for the conduct of scoping reviews", *JBI Evidence Synthesis*, 18(10), pp. 2119–2126, 2020.
- [16] Hong Q. N., Pluye P., Fàbregues S., Bartlett G., Boardman F., Cargo M., Dagenais P., Gagnon M.-P., Griffiths F., Nicolau B., O'Cathain A., Rousseau M.-C., Vedel I., "Mixed Methods Appraisal Tool (MMAT), version 2018", Canadian Intellectual Property Office, 2018.
- [17] Thomas J., Harden A., "Methods for the thematic synthesis of qualitative research in systematic reviews", *BMC Medical Research Methodology*, 8, Article 45, 2008.
- \*[18] Yani A., Ahmad S. S. B. H., Jaili H. H. B. H., Karim H. N. B. H. A., Abed E. M. H., "Arabic language competencies for academic and professional purposes in generative AI-enhanced learning and workplace environments", *International Journal of Artificial Intelligence and Machine Learning*, 6(7S), pp. 79–90, 2026.
- \*[19] Oh H. J., "Developing hybrid expertise through generative AI-mediated ESP course design: A three-year longitudinal case study of a Bio-health English course", *English Teaching*, 81(1), pp. 105–125, 2026.
- \*[20] Altay M., Aykut Kolay C., Yöney R. Z., "Pre-service EFL teachers' EAP literacies: A transition to GenAI-based curriculum design", *Journal of English for Academic Purposes*, 82, Article 101710, 2026.
- \*[21] Rutecka P., Cicha K., Rizun M., Strzelecki A., "Generative AI in curriculum design: Empirical insights into model performance and educational constraints", *IEEE Transactions on Learning Technologies*, 18, pp. 757–768, 2025.
- \*[22] Liu J. E., Qiu Z., Lin A. M. Y., "Integrating GenAI with genre-based pedagogy and the PAA model in a college writing course: Enhancing critical engagement in an AI-mediated learning ecology", *Linguistics and Education*, 92, Article 101487, 2026.
- \*[23] Ngo T. N., Hastie D., "Artificial Intelligence for Academic Purposes (AIAP): Integrating AI literacy into an EAP module", *English for Specific Purposes*, 77, pp. 20–38, 2025.
- \*[24] Ting A., "Incorporating AI literacy into a university professional communication English class", *Journal of English for Academic Purposes*, 2026.
- \*[25] Alghamdi M., Ahmed S. N., "Evaluation of an AI-integrated English communication curriculum intervention in Saudi Arabia", *Discover Education*, 5(1), Article 532, 2026.
- \*[26] Singh R. G., Chow K. M., Dong H., Wong H., Fung M., Myers P. S., Wong L. L. C., Lai P., Ngai C. S. B., Koon A. C., Yuen B., Wong J., Jhaveri A., "Enhancing academic English instruction through pedagogically integrated GenAI: Effects on students' AI literacy and engagement", *Journal of English for Academic Purposes*, 2026.
- \*[27] Boustane H., Sabil A., Chroqui R., "Developing an online course to enhance AI-assisted academic writing among engineering students: A design-based research study", *Lecture Notes in Networks and Systems*, 1878, pp. 205–212, 2026.



- \*[28] Dou W., "The application of generative AI technology in ESP courses for civil engineering majors", Proceedings of ICAIDT 2024, 2024.
- \*[29] Wang J., "Large-language-model-supported BOPPPS teaching for maritime ESP: A quasi-experimental study in a Chinese vocational college", Proceedings of ICIEAI 2025, 2026.
- \*[30] Chen Q., Tawilapakul U., Lin A. M. Y., "Teacherness in the age of GenAI: An ethnographic exploration of pedagogical decision-making and AI integration in English language teaching", System, 136, Article 103877, 2026.
- \*[31] Ngo T. N., Fians G., "Between prompting strategies and ethical concerns: How EAP teachers develop materials with generative AI", Journal of English for Academic Purposes, 81, Article 101680, 2026.
- \*[32] Crosthwaite P., Özer M., "Vibecoding as method: A structured workflow for AI-assisted research instrument development in applied linguistics", Research Methods in Applied Linguistics, 2026.
- \*[33] Ngo T. N., "Navigating ESAP materials development with generative AI support", Innovation in Language Learning and Teaching, 2026.
- \*[34] Kohnke L., Zou D., "Developing GenAI literacy in EAP teacher education: A case study of materials design and professional identity", Journal of English for Academic Purposes, 2026.
- \*[35] Ma X., Jiang F., "From general academic to disciplinary literacy: A collaborative autoethnography of GenAI-enhanced EAP materials development", Journal of English for Academic Purposes, 2026.
- \*[36] Gao X., "From unified to differentiated materials: Generative AI-supported adaptation of EAP reading materials", Frontiers in Psychology, 17, Article 1887565, 2026.
- \*[37] Huang X., Zhao H., Dang T. N. Y., "Exploring the potential of ChatGPT to create reading materials for incidental learning of academic vocabulary", Journal of English for Academic Purposes, 84, Article 101777, 2026.
- \*[38] Nikolova-Stoupak I., Bibauw S., Dumont A., Watrin P., François T., "Generating contexts for ESP vocabulary exercises with LLMs", Proceedings of the 13th NLP4CALL Workshop, 2024.
- \*[39] Ellederova E., "Integrating AI into English for electrical engineering and information technology learning materials", International Conference the Future of Education, 2025.
- \*[40] Fu X., Zhang L. J., "Integrating GenAI into EAP course materials design: A case study", Journal of English for Academic Purposes, 2026.
- \*[41] Yu D., "Towards LLM-assisted move annotation: Leveraging ChatGPT-4 to analyse the genre structure of CEO statements in corporate social responsibility reports", English for Specific Purposes, 2025.
- \*[42] Kim M., Lu X., "Exploring the potential of using ChatGPT for rhetorical move-step analysis", Journal of English for Academic Purposes, 2024.
- \*[43] Sfeir M., "Do-it-yourself corpora and discipline-specific writing: A focus on the benefits of corpus building and cleaning (in the age of large language models)", English for Specific Purposes, 2025.
- \*[44] Dong Y., Wang S., Li Y., "Enhancing human-AI communication in language education: Developing customized chatbots for teaching genre analysis and annotation", Linguistics and Education, 2026.
- \*[45] Warden C. A., Stanworth J. O., Chen J. F., "Generative AI and the emergence of hybrid authorship in ESP writing: A corpus-based analysis (2015–2024)", English for Specific Purposes, 84, pp. 244–259, 2026.
- \*[46] You X., You X., "AI-mediated genre learning in the business English class", English for Specific Purposes, 2026.
- \*[47] Korzin A. S., "Using AI and corpus-based tools for teaching terminological collocation in the field of aerospace engineering", Proceedings of ICAITE 2025, 2025.
- \*[48] Elmotri B., Osman W. A. E. A. E., Al-Humari M. A., Garrouri S., Hassanein O. H. S., "Possibilities for improving ESP curriculum design and assessment strategies for Saudi universities with ChatGPT", Arab World English Journal, 2025(Special Issue), pp. 234–250, 2025.
- \*[49] Lin R., Chen J., Deng F., "Exploring teacher resilience in developing EAP reading assessment materials with Gen-AI", Journal of English for Academic Purposes, 83, Article 101757, 2026.
- \*[50] Urquhart L., Ngo X. M., "Changing EAP assessment practices in the age of generative artificial intelligence: The case of Scottish higher education institutions", Journal of English for Academic Purposes, 79, Article 101609, 2026.
- \*[51] Kovačević D., "ChatGPT and creation of tests in ESP", Proceedings of the 24th International Symposium INFOTEH-JAHORINA, 2025.
- \*[52] Aryadoust V., Wong J., Zhang W., Qian J., Sai Z., Liu T., Han Y., "Automatic conversational assessment using a GPT-based model for EAP speaking", English for Specific Purposes, 84, 2026.



\*[53] Assassi T., “ESP practitioners’ perspectives on the efficiency of using artificial intelligence ‘ChatGPT’ in needs analysis”, *Computer Assisted Language Learning*, 39(4), pp. 797–828, 2026.

\*[54] Das M., Li X., Fabunmi O., Hölttä-Otto K., “User needs analysis in design research by using large language models as interviewees”, *Journal of Computing and Information Science in Engineering*, 26(8), Article 081002, 2026.

\*[55] Barandoni S., Cascone L., Chiarello F., “Automating customer needs analysis: A comparative study of large language models in the travel industry”, *Computers in Industry*, 2026.

\*[56] Boustane H., Chroqui R., Sabil A., “Bridging gaps in generative AI use for academic writing: Insights from Moroccan engineering students”, *Lecture Notes in Networks and Systems*, 1640, pp. 251–262, 2026.

[57] Yin S., Chapelle C. A., “A systematic examination of generative artificial intelligence (GenAI) use guidelines in applied linguistics journals”, *Research Methods in Applied Linguistics*, 2025.

[58] Argyle L. P., Busby E. C., Fulda N., Gubler J. R., Rytting C., Wingate D., “Out of one, many: Using language models to simulate human samples”, *Political Analysis*, 31(3), pp. 337–351, 2023.

[59] Bisbee J., Clinton J. D., Dorff C., Kenkel B., Larson J. M., “Synthetic replacements for human survey data? The perils of large language models”, *Political Analysis*, 32(4), pp. 401–416, 2024.