



Mapping the Field of Peer Review Reports in Applied Linguistics: A Scoping Review

Sahar Hayatolain¹ , Mansoor Ganji^{2*} , Nahid Yarahmadzahi¹ ,
Amir Mohammadian¹ 

¹ Candidate in TEFL, English Department, Faculty of Management and Humanities,
Chabahar Maritime University, Chabahar, Iran

² Department of English Language and Literature, Faculty of Humanities and Social
Sciences, Golestan University, Gorgan, Iran

Received: 2026/04/18

Accepted: 2026/06/01

Abstract: In the current competitive academic world, publishing papers in prestigious journals is of great significance to scholars, and one of the main elements in ensuring the quality of submitted manuscripts is peer review. The fate of manuscripts largely depends on the final review reports provided by the reviewers. Research has investigated peer review from different aspects. However, a comprehensive account of all the studies in this regard, especially within the field of applied linguistics and focusing on authentic blind peer review reports, is missing. Accordingly, this scoping study aimed to provide a clear picture of existing research on peer review in applied linguistics. To this end, following Arksey and O'Malley's (2005) scoping review guidelines, a total of 33 studies were examined to assess their scope, addressed themes, and methodological trends. The results revealed that these studies were published between 2004 and 2024, in 27 outlets, originated from 15 countries, and mainly focused on single-journal contexts. The main themes investigated in these studies were comparative analyses, pragmatics, comment focus, genre structure, metadiscourse, and association, with comparative analyses being the most frequent. In terms of methodological traits, most studies were qualitative and descriptive. This review can benefit authors, reviewers, and editors by clarifying the current status of research on peer review and providing a foundation for future research on peer review practices.

Keywords: Applied Linguistics, Peer Review, Publishing Papers, Scoping Review.

* Corresponding Author.

Authors' Email Address:

¹ Sahar Hayatolain (saharhayat1996@gmail.com), ² Mansoor Ganji (ganjimansoor1980@gmail.com), ³ Nahid Yarahmadzahi (venayarahmadi@gmail.com), ⁴ Amir Mohammadian (amiancmu@gmail.com)



Introduction

A key component of scholarly communication is peer review, essential for maintaining scientific integrity, correcting errors, and ensuring the publication of the highest quality research (Aczel et al., 2025; Drozd & Ladomery, 2024). Essentially, reviewers perform the role of gatekeepers of scientific knowledge (Sizo et al., 2025), provide constructive feedback, and help editors in rejecting or accepting a manuscript (Kelly et al., 2014; Tumin & Tobias, 2019). Their contribution is very important since scholarly output has grown at a noticeably rapid rate (Drozd & Ladomery, 2024), and researchers are facing the increasing pressure to publish in various fields (Lu, 2022). The frequency of manuscript submissions is on the rise annually (by 6.1%) and retractions are increasing each year since 2013 (Publons, 2018). However, the requirements in submission and review may vary considerably from field to field.

It is estimated that 15 million hours per year are spent reviewing previously rejected and resubmitted manuscripts (American Journal Experts, 2018), resulting in reviewer fatigue and uneven coverage of submissions (Drozd & Ladomery, 2024). Moreover, the focus on metrics has led to negative outcomes like plagiarism or manipulation of data (Bahl & Bahl, 2021; Paruzel-Czachura et al., 2021); accordingly, poor-quality journals gain from the rush for fast publishing by lowering their standards for editing and peer review (Marina & Sterligov, 2021). As a result of these trends, the review system may not reliably safeguard reproducible science.

Practitioners recognize that traditional peer-review practices are ineffective, plagued with long delays, different quality standards, and inconsistent results (Ross-Hellauer et al., 2023). Meta-analytic evidence shows substantial variation in how the same manuscript might be judged by different reviewers (Bornmann et al., 2010). Consequently, major manuscript errors are often overlooked, and reviewers miss errors linked to irreproducible results (Aczel et al., 2025). In the same vein, Ganji and Derakhshan (2020) state that this practice hinders original ideas, wastes time and financial resources, and remains highly subjective; furthermore, it often fails to identify the demerits of research. In addition, biomedical researchers often blame poor-quality peer review for irreproducibility (Cobey et al., 2024). The unreliability may be increased since most journals do not provide reviewers with quality guidelines or training (Song et al., 2021). As a result, reports can become idiosyncratic and biased (Bornmann & Mungra, 2011), especially in the social sciences and second round reviews (Aczel et al., 2025).

The effects of this inconsistency are more severe when it comes to the communicative aspect of the review process. Due to the very nature of applied linguistics, where the focus is on language per se, the language of feedback is of great importance. The reviewers should strike a balance between being critical and developmental; use language that is honest, respectful, and straightforward instead of being focused on insignificant details (Ellinger et al., 2013). Critical comments should be polite and properly structured, with proper hedging and specifics. On the other hand, unclear or culturally unrelated feedback may complicate the revision process. For example, responding to the reviewers' comments was one of the hardest problems faced by non-native English writers (Mudrak, 2013; Sizo et al., 2025).

However, a comprehensive understanding of the feedback's communicative function has not yet been reached. In applied linguistics, the discourse of peer review has been investigated in several separate studies, but their results remain confined to small sub-areas. For example, such patterns as summarizing a judgment or overall opinion, offering critical comments, and providing recommendations have been identified (Kheradparvar & Dreyfus, 2025; Mason & Chong, 2022; Yakhontova, 2019). Besides, some researchers have focused on the evaluative and interpersonal aspects of review language (e.g., Hewings, 2004), while there are also studies focusing on the appraisal mechanisms (Kheradparvar & Dreyfus, 2025) and politeness strategies in order to dilute criticism and create an egalitarian tone (e.g., Sönmez & Akbas, 2023), as well as those highlighting interactional aspects in reviews (e.g., Gosden, 2003). Empirical studies have addressed the discrepancy between content-based and language-based feedback, the tendency to favor lexical and syntactic elements (e.g., Belcher, 2007; Kashiha, 2024), and the perceptions of the reviewers regarding their agency (e.g., Mohammadi et al., 2023).

Literature Review

The study of peer review reports and comments has become a popular trend in recent years (Kashiha, 2024), partly because the role of reviewers is vital in the creation of academic communication and the attribution of quality in the academic environment (Stupacher, 2025). Besides, a review is a complex social genre characterized by complex linguistic and rhetorical conventions. Recent studies in different fields have tried to unveil these conventions, investigating the impact of reviewer characteristics and final assessment on the given feedback (Atjonen, 2019; Buljan et al., 2020; Falk Delgado et al., 2019; Garcia-Costa et al., 2022).

As regards the general structure of the review reports, there were several studies mostly done in the field of applied linguistics. To describe the organization and key linguistic features of peer review reports, [Kheradparvar and Dreyfus \(2025\)](#) analyzed 18 accepted with revision peer review reports from applied linguistics journals. They identified the two-stage pattern of overview followed by recommendations for revisions in the reports, similar to previous studies on reports from different fields. Similarly, in [Gosden's \(2003\)](#) analysis of peer review reports, the structure comprised a preliminary summarizing judgment regarding the manuscript's fitness for publishing, followed by the reviewer's comments. This structure was also observed in [Fortanet-Gomez \(2009\)](#) and [Paltridge \(2017\)](#). [Kheradparvar and Dreyfus \(2025\)](#) also showed that reviewers open with explicit praise but deliver most criticism through implicitly negative evaluative language (e.g., modalized "should..." comments). In the end, the results of [Kheradparvar and Dreyfus \(2025\)](#) showed that peer review in applied linguistics, although still characterized by the same general structuring as in other scientific domains, has some peculiarities as far as the language of the reviewers and the feedback strategies are concerned, which are specific to the discipline, and even classified accordingly.

[Paltridge \(2015\)](#) found that reviewers often frame their requests indirectly, using phrases such as "I suggest", which might cause some difficulty for novice authors in interpreting them. In this way, the current evidence regarding the indirectness characteristic of review comments opens up a new area for research. [Yakhontova \(2019\)](#) worked with 34 double-blind reviews (17 from each field) and identified a three-move structure consisting of summarizing the reviewer's overall opinion, providing critical comments and recommendations, and, optionally, giving a final recommendation in the corpus. Furthermore, applied linguistics reviews were found to be more interpersonal, less formal, and more reliant on recommendations and suggestions than their mathematical counterparts. In the same vein, analyzing 160 double-blind review reports on Iranian submissions to applied linguistics and engineering journals, [Kashiha \(2024\)](#) found that in both fields, reviewers' comments were more language-oriented rather than being content-oriented, followed a major to minor or section by section issue identification pattern, and started with a synopsis which combined initial praise with subsequent criticism in a "positive news first, negative news later" pattern.

Apart from empirical studies, as far as scoping reviews are concerned, a number of studies have identified various dimensions of peer review in different disciplines. Among these are scoping reviews of reviewers' roles and responsibilities in biomedical journals ([Glonti et al., 2017](#)), investigations into peer review guidelines and checklists used in biomedical publishing ([Song et al., 2021](#)), an empirical synthesis of research on open peer

review practices and outcomes (Ross-Hellauer & Horbach, 2024), and mappings of criteria and quality indicators for effective peer review reports across disciplines (Ganji & Derakhshan, 2020; Sizo et al., 2025). However, despite the growing body of research on peer review discourse, no comprehensive scoping review has yet synthesized studies specifically examining peer review comments within applied linguistics.

As a result of this gap, a systematic mapping of research on peer review reports is needed to integrate scattered findings and provide an evidence-based overview of prevailing trends, themes, and methodologies. A scoping review is uniquely appropriate for this study rather than a systematic review. While a systematic review is designed to answer a narrow, specific question often focused on the clinical efficacy of an intervention, a scoping review is meant to broadly map the extent, range, and nature of research activity within an area that contains heterogeneous and scattered literature (Peters et al., 2020). Hence, the current study aims to answer the following three research questions:

1. What is the extent and distribution of research on peer review comments in applied linguistics with respect to publication profiles (years, types, and venues), geographical regions, study contexts, target languages, and data collection periods?
2. What are the main themes examined in research on peer review comments in applied linguistics?
3. What research paradigms, research designs, data collection methods, and analytical techniques are employed in the studies of peer review comments in applied linguistics journals?

Methodology

Design

In this study, a scoping review design was chosen to map and describe the existing research on peer review comments in applied linguistics. In doing so, researchers are able to investigate how a topic is conceptualized, track methodological developments, and uncover knowledge gaps without a need to combine findings to reach a single conclusion (Hillman et al., 2021; Peters et al., 2020).

Conforming to Joanna Briggs Institute (JBI) guidelines (Peters et al., 2020), a formal quality appraisal or risk of bias assessment was not conducted. Rather than identifying best practices, this review aimed to map the wide range of methodological and conceptual choices currently used regarding peer review comments. Therefore, evaluating the methodological rigor of individual studies fell outside the scope of this review.

An a priori protocol guided this scoping review. It is worth mentioning that, given the small and exploratory nature of the study, the research team decided that formal registration on platforms such as OSF or Figshare was not performed.

Frameworks

The review protocol was conducted based on the methodological guidelines for scoping reviews published by the Joanna Briggs Institute (JBI) ([Peters et al., 2020](#)). In addition, steps for conducting a scoping review outlined by [Arksey and O'Malley \(2005\)](#) were followed in the study. Figure 1 illustrates the stages of the review:

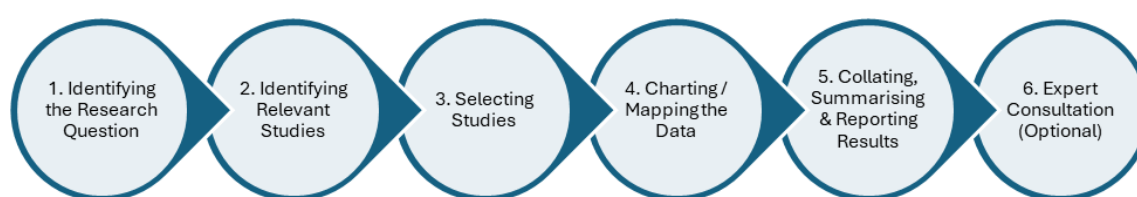


Figure 1. [Arksey and O'Malley \(2005\)](#) Methodological Stages for Conducting a Scoping Review (p. 22)

In this review, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) are used to ensure transparent and complete reporting ([Tricco et al., 2018](#)). PRISMA-ScR is used as a reporting guideline rather than a methodological framework, as it specifies what information should be reported when presenting scoping review methods and results. In particular, PRISMA-ScR is applied in this review to Stage 3 and the final Results section, where the study selection process will be summarized using the PRISMA-ScR flow diagram.

Instruments

Several instruments were utilized in this study to ensure the systematic and reproducible process of conducting the research. The SPIDER tool by [Cooke et al. \(2012\)](#) was used to develop search strings that would lead to the discovery of relevant studies, while Zotero software was used to manage, deduplicate, and screen the selected articles. Other research instruments that were utilized included a data-charting form ([Appendix C](#)) and thematic tables that were developed within Microsoft Word. Detailed explanations of how and where these tools were applied are provided in the subsequent stages of the methodology.

Scoping Review Procedure

Stage 1: Identifying the Research Questions

Following JBI guidance, three research questions were defined with the aim of mapping the scope and nature of research on peer review comments in applied linguistics. The questions are as follows:

Mapping Question: Research question one on extent and distribution

Theme Question: Research question two on main themes

Methodological Question: Research question three on methodology

Consistent with the purpose of scoping reviews, these questions were designed to support descriptive mapping, conceptual clarification, and methodological profiling.

Stage 2: Identifying Relevant Studies

Two major multidisciplinary and discipline-specific electronic databases were included in this search (Scopus, Web of Science) to provide the widest possible coverage of the literature.

Search Strategy

The search strategy followed the JBI three-step approach. The first step included an initial exploratory search, which used Scopus and Web of Science to find appropriate keywords and index terms. The second step involved conducting a complete database search through selected databases with the use of improved search terms and controlled vocabulary. Lastly, manual screening of the reference lists of all the included studies was conducted to identify any possible additional studies. Search engines such as Google Scholar and ResearchGate were also occasionally used to access the full-text version of the paper.

As mentioned earlier, considering the qualitative, discourse-analytic nature of peer review comment research, the search terms were informed by the SPIDER tool (Cooke et al., 2012). This tool facilitates the sensitive retrieval of qualitative and mixed-methods studies. A complete search was conducted using search terms that were created according to SPIDER criteria to examine the sample, phenomenon of interest, design, evaluation, and research type. The sample aspect included peer review texts and Applied Linguistics professionals using terms "applied linguistics", "TESOL", "EFL", and "peer reviewers". For the phenomenon of interest, the focus was on peer-review comments that evaluate and guide revisions, using terms such as "peer review," "reviewer feedback," and "manuscript evaluation." The search included different study designs, such as discourse analysis and mixed methods, and examined features like stance, tone, and bias. The researchers used Boolean operators

(AND/OR) to show how these elements connect to one another. The researchers maintained an unbiased approach to research types, which allowed them to include all empirical studies that used qualitative, quantitative, and mixed-methods research approaches. It is noteworthy that each search string was adapted to the indexing syntax of the respective database. Searches were conducted from 1 to 20 November 2025, and results were restricted to English-language publications. It is noteworthy that adjustments were made to ensure compatibility with the specific search syntax of the databases. Shorter and alternative variations of the search terms were also used in each database to capture additional relevant studies. The search strings are outlined in Appendix B.

Eligibility Criteria (PCC Framework)

Eligibility criteria were developed according to the Population–Concept–Context (PCC) framework recommended for scoping reviews by JBI (Peters et al., 2020). Specifically, we included journal articles and book chapters that reported empirical analyses of peer review comments. Only studies with authors affiliated with applied linguistics or related language-education fields were considered. Regarding the focus of the studies, included studies had to analyze actual peer review comments, rather than other aspects such as opinions on the review process. The studies were selected only if they examined some aspects of the comments themselves. In terms of document characteristics, we focused on English-language publications in English-medium applied linguistics journals. Furthermore, publications were limited to studies published between 1 January 2000 and 20 December 2025. The details of these criteria are presented in Table 1.

Table 1. Eligibility Criteria of the Study Based on the PCC Framework

Criterion	Include	Exclude	Rationale
Population	Journal papers, book chapters, peer reviewers, and authors.	Student peer review in classrooms, grant proposals.	Specifies the exact targets of the review process.
Concept	Empirical analyses of actual peer review comments on manuscripts	Open peer reviews, theoretical discussions, editorials, non-empirical research, and literature reviews.	Ensures focus on evidence-based insights
Context	Peer-reviewed academic publishing (journal articles and scholarly book chapters) within applied linguistics and its subfields, published in English between 1 January 2000 and 20 December 2025	Publications outside applied linguistics, conference papers, dissertations, grey literature, non-English manuscripts, and studies published outside the date range.	Ensures disciplinary relevance, maintains comparability of high-quality sources, and captures contemporary editorial developments.

Stage 3: Selecting Studies

All the retrieved records were exported to Zotero reference management software. A total of 114 studies from Scopus, 106 from Web of Science, and 33 from manual searching and snowballing were found. All duplicates were identified and removed ($n=50$). Deduplicated results (203 studies) were then imported into a separate Zotero library to manage the screening process.

Screening was conducted in two steps: (1) title and abstract screening, and (2) full-text screening. Each stage was performed independently by two reviewers. In the first stage, the reviewers skimmed the titles and abstracts of the studies to decide whether they should be included for full-text screening. After excluding 157 studies, 46 studies were selected for retrieval, one of which was not retrieved because it was only available in print and not accessible online.

In the second stage, after full-text screening of 45 studies, 12 were excluded due to a different focus, field, language, corpus, or non-empirical design, not being retrievable, or being a book chapter of the same paper. Finally, 33 studies were chosen as the final corpus to be included in the scoping review.

In the event of a disagreement, it was resolved by consensus; if necessary, a third reviewer arbitrated to ensure consistency. Reasons for exclusion at the full-text screening stage were documented for full transparency. The selection process is summarized in the following PRISMA-ScR flow diagram, showing the number of records identified, screened, excluded, and included.

Materials

The final materials of the study, as suggested in the PRISMA-ScR diagram, included 33 empirical studies. They were either journal articles or chapters in books, which formed the basis of the analyses in the next stages. Full details of the included studies are available in Appendix A.

Stage 4: Data Charting

A data charting form was constructed on Microsoft Word by the researchers following JBI (Peters et al., 2020). The purpose of this form was to help us consistently and clearly collect the same information from all the studies we included. The researchers created data-charting categories by using both the common patterns found in the included studies and the research inquiries established for the review. In doing so, the trends in publications, the main themes,

and the research methods used were more convenient to identify and document. Table 2 displays the variables we looked for in the selected papers.

The final charting form is available in Appendix C.

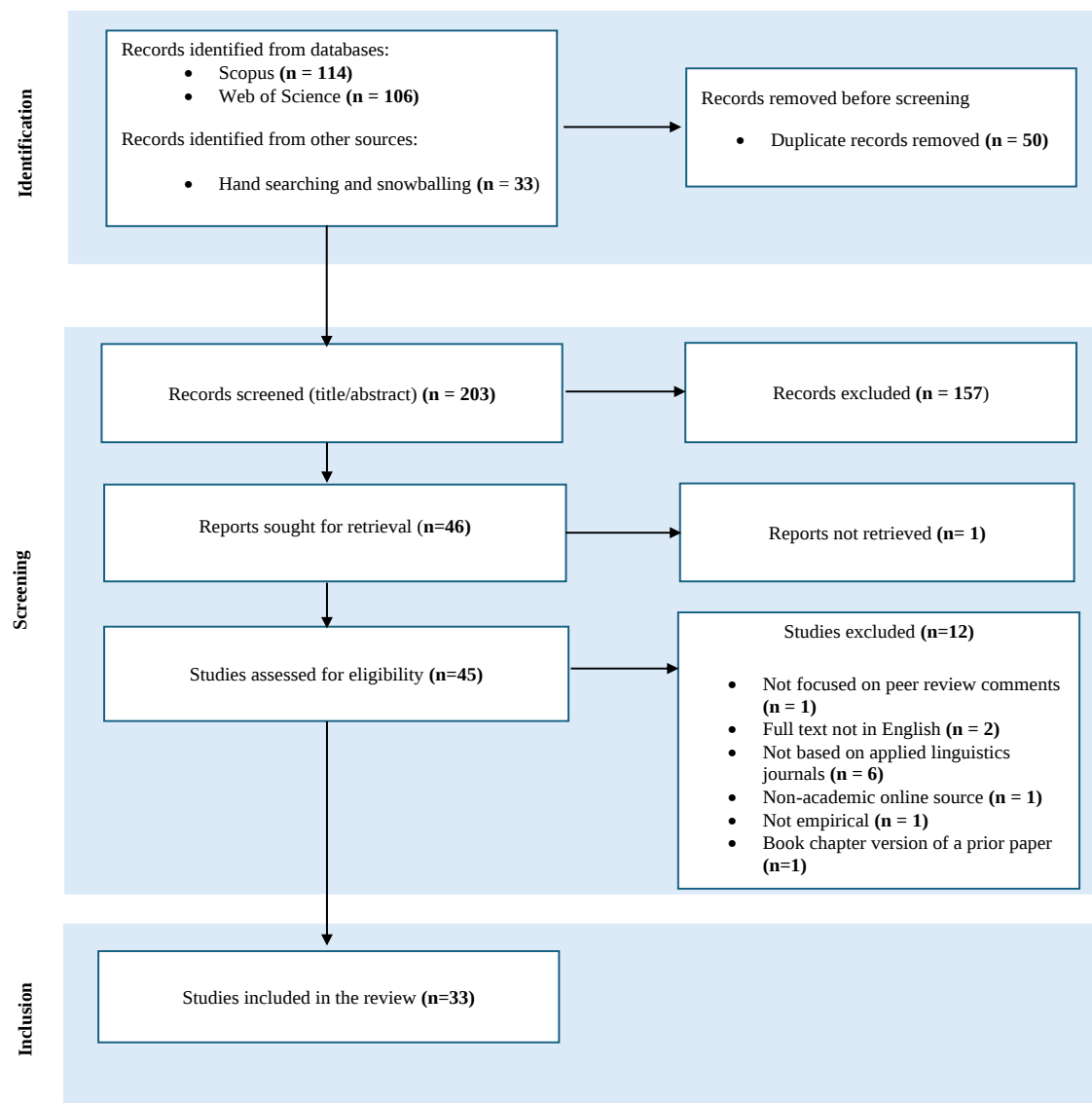


Figure 2. PRISMA-ScR Screening Process

Table 2. Data-Charting Scheme of the Study

Category	Variables
Bibliographic	Author, Year, Journal/Source, Publication type
Context	Country (Affiliation & Study), Region, Target language
Aim	Research purpose / Objectives
Sample / Data	Sample size (number of reviews), Data collection period
Thematic Focus	Primary themes, Key variables/constructs
Method	Research design, Analysis type
Framework	Theoretical or Analytical model

To make sure that the form was able to capture the necessary details about peer review comments, it was pilot-tested in five of the included studies. Each of the two reviewers assessed the included studies independently. Inter-rater agreement was calculated via Cohen's kappa, which yielded a value of 0.87, indicating strong agreement levels. Next, the team compared the charted entries to identify differences, which they resolved through discussion until they reached agreement on uniformity. The research team conducted multiple meetings to refine the charting categories and decision rules until they achieved clearer and more reproducible assessment methods.

For the first and third research questions, a data charting approach was used. Descriptive information, including publication year, journal, author affiliation, study context, target language, and data collection period (RQ1), as well as research paradigm, design, data sources, and sample size (RQ3), was extracted directly from the texts of the included studies. Since the aim was to provide a descriptive account of these study characteristics, no interpretive coding was applied. On the flip side, RQ2 involved an inductive thematic coding process. The specific analytical focuses of each study (e.g., “positive/negative expressions”) were extracted and then assigned to emergent themes (e.g., Pragmatics, Metadiscourse).

Stage 5: Collating, Summarizing, and Reporting the Results

In this stage, the charted data were organized and summarized to address the research questions of the review. Descriptive and qualitative methods were used to present an overview of the extent, thematic focus, findings, and methodological characteristics of studies on peer review comments in applied linguistics.

To answer the first research question, the research team compiled descriptive statistics, which included publication year, journal, country, target language, and sample size, using tables and figures to display the resulting trends and distributions. Regarding the second research question, the emergent themes identified in inductive thematic analysis were presented to show the main thematic orientations of the studies. Frequency counts of themes and their codes were calculated and illustrated in figures. Finally, to answer the third research question, we summarized the methodological details from each article included in the study. Information on research paradigms, study designs, data collection methods, and analysis techniques was then organized into charts and tables to help show the overall methodological landscape of the field.

Results

Research Question One: Extent and Distribution of Included Studies, Including Publication Profiles, Geographical Coverage, and Study Contexts

The first research question aimed at describing the included studies in terms of publication profiles, geographical coverage, study contexts, target languages, and data collection periods. The reviewed studies span more than two decades (2004–2025), with the earliest publication appearing in 2004 (S9: [Hewings, 2004](#)). Only a limited number of studies ($n = 5$; S2, S7, S8, S9, S10) were published before 2010. From 2015 onwards, the volume of research increased substantially, with $n = 24$ studies published between 2015 and 2025, including a noticeable concentration in 2020–2025 ($n = 15$). Figure 3 displays the change in the number of publications from 2004 to 2025.

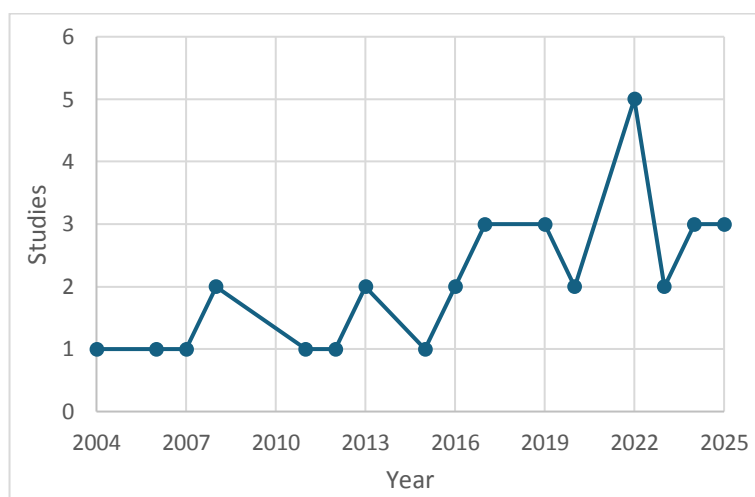


Figure 3. Number of Included Studies by Year (2004–2025)

Most studies were published as journal articles ($n = 23$; 69.7%), while a smaller proportion appeared as book chapters in edited volumes and monographs ($n = 10$; 30.3%). Figure 4 shows these proportions.

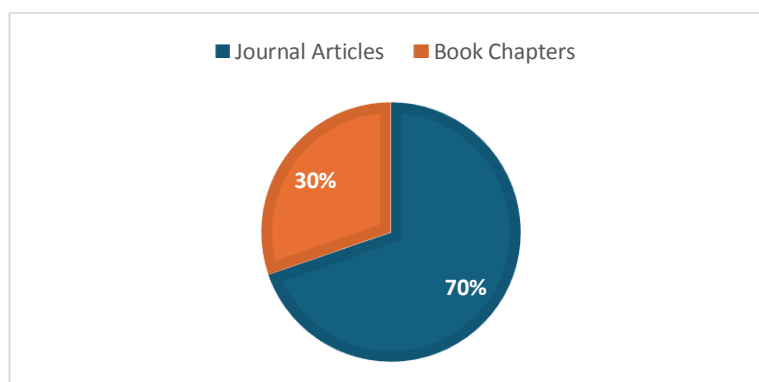


Figure 4. The Proportion of Publication Types

As regards publication venues, the most frequently represented journals were *Ibérica* (n=2; S14, S33), *System* (n=2; S4, S5), and *Journal of English for Academic Purposes* (n=2; S7, S12). Other journals appeared once each, including *English for Specific Purposes* (n=1; S27), *Studies in Higher Education* (n=1; S21), and *Text & Talk* (n=1; S13). Several chapters in the books “*The Discourse of Peer Review*” (n=3; S22–S24) and “*The Inner World of Gatekeeping in Scholarly Publication*” (n=2; S17, S28) served as recurring venues for book-chapter publications in the corpus. All in all, there was a total of 27 distinct venues for the 33 studies.

The geographical distribution of the included studies was examined in terms of the first authors’ institutional affiliation country. There was a total of 15 countries represented by the affiliations. The largest shares belong to Australia (n = 7; 21.2%; S16, S21–S26) and Oman (n = 5; 15.2%; S11–S15), mainly due to the significant contributions by Paltridge and Kashiha in those countries, respectively. Next most frequently represented countries were Spain (n = 3; 9.1%; S3, S7, S8) and the United Kingdom (n = 3; 9.1%, S9, S10, S29). The United States (S2, S27), Ukraine (S31, S32), Russia (S18, S19), and Hong Kong (S4, S5) accounted for a small portion of research, with each one contributing two studies (6.1% each). The remaining countries were represented by single studies (n = 1; 3% each): Canada (S28), Iran (S33), Japan (S20), Saudi Arabia (S1), Sweden (S17), Thailand (S30), and the United Arab Emirates (S6).

As far as the data sources of the studies are concerned, they were categorized as is evident in Table 3.

Table 3. Distribution of Study Contexts (Data Sources)

Category	Subcategory / Details	Study IDs	Frequency (n)	Percentage (%)
Single Journal Contexts	<i>English for Specific Purposes</i> (ESPJ)	S2, S9, S10, S21, S22, S23, S24, S25, S26, S27	10	30.3%
	<i>System</i>	S4, S5	2	6.1%
	<i>Russian Journal of Linguistics</i>	S18, S19	2	6.1%
	<i>Ibérica</i>	S3	1	3.0%
	<i>ReCALL</i>	S29	1	3.0%
Specified Multi-Journal	Datasets listing specific journals or publishers	S11, S12, S13, S14, S15, S16, S17, S33	8	24.2%
General / Unspecified International	Datasets from multiple journals without specific names listed	S1, S6, S7, S8, S20, S28, S30, S31, S32	9	27.3%
Total		S1–S33	33	100%

Totally, 48.5% (n=16) of studies focused on a single publication outlet. The journal *English for Specific Purposes* (ESPJ) served as the most prominent context, accounting for 30.3% of the total research corpus. However, 24.2% of studies mentioned the specific multi-journal databases, and 27.3% obtained their data from international or regional sources without disclosing the journal names.

English served as the main target language for the majority of the studies (n=31; 93.9%), with only two studies examining a bilingual English–Russian corpus (S18, S19; 6.1%). Regarding temporal coverage (i.e., the timeframe from which the analyzed peer-review reports were drawn), reporting was inconsistent across the included studies. An explicit data collection period was provided in half of the studies (n = 17; 51.5%; S1–S5, S9–S10, S15–S18, S20, S27–S29, S32–S33), specifying either a particular year or a clearly defined date range. Six studies referred to a one-year period without stating the exact year (n = 6; 18.2%; S21–S26). Nine studies did not report a clear data collection period (n = 9; 26.5%; S6–S8, S11–S14, S19, S31), and one additional study provided only an imprecise timeframe (“last few years before 2013”; S30, 3.0%). Details are presented in Table 4.

Table 4. Temporal Coverage (Data Collection Periods) Reported in Included Studies (N = 33)

Reporting category	n	%	Study IDs	Temporal coverage reported (exact wording/years)
Explicit year/range reported	17	51.5	S1	2016–2020
			S2	1998–2001
			S3	1999–2013
			S4, S5	2003–2011
			S9	Dec 2001 – Apr 2004
			S10	Dec 1998 – Apr 2004
			S15	2015–2020
			S16	2014–2020
			S17	2012–2022 (over the last decade)
			S18	2014–2019
			S20	2017–2021
			S27	2007
			S28	2010–2019
			S29	2015–2022
			S32	2006–2022
			S33	2009–2014
One-year period (year not specified; stated or inferred)	6	18.2	S21–S26	One-year period (calendar year not specified)
Imprecise timeframe	1	3.0	S30	Last few years before 2013 (specific year not mentioned)
Not reported / not specified	9	27.3	S6, S7, S8, S11, S12, S13, S14, S19, S31	Not reported / not specified
Total	33	100.0	—	—

Research Question Two: Thematic Focus of Included Studies

To answer the second research question, which focused on the main themes explored in the selected studies, a thematic analysis of the 33 studies was conducted. As a result, six main thematic categories, including Comparative analyses, Pragmatics, Comment focus (targets of critique, language vs content), Genre structure, Metadiscourse (stance/engagement), and Association emerged. The operational definitions and some details on the inclusion and exclusion criteria, and some examples from the included studies are provided in Table 5.

Table 5. Thematic Classification of Peer Review Studies: Definitions, Criteria, and Examples

Theme	Operational Definition	Inclusion Criteria	Exclusion Criteria	Example from Dataset
Comparative	Explores the ways reviewers' comments vary in different situations. It considers the effects of editorial outcomes, discipline of study, reviewer backgrounds, cultural settings, and review methods.	Studies systematically comparing comments across two or more distinct variables (e.g., editorial outcomes, disciplines, author backgrounds).	Studies analyzing comments from a single, uniform group without any cross-group comparison.	Comparing the discourse moves in accepted vs. rejected manuscripts reviews (e.g., S1) or between Anglophone vs. non-Anglophone authors (e.g., S3).
Pragmatics	Investigates how reviewers use language when evaluating manuscripts to balance their roles as advisors and evaluators using politeness	Studies focusing on speech acts, politeness/impoliteness, mitigation, tone, or how reviewers perform identity roles (e.g., gatekeeper vs. mentor).	Studies analyzing only content or genre without considering interpersonal strategies.	Analyzing specific speech acts like directions, suggestions, or indirect requests (e.g., S21), or cultural

Theme	Operational Definition	Inclusion Criteria	Exclusion Criteria	Example from Dataset
	strategies, directives, persuasive language, and role positioning.			differences in mitigation (e.g., S18).
Comment Focus	Takes into account the ways reviewers classify their feedback to address content or the language of the manuscripts. It also focuses on the distribution of comments in different sections of the manuscripts.	Studies categorizing what the feedback is about (e.g., content vs. language, methodology) or where it is located in the manuscript.	Studies analyzing <i>how</i> the critique is delivered rhetorically rather than <i>what</i> specific manuscript flaws it targets.	Categorizing linguistic infelicities (e.g., S3) or classifying comments by manuscript section and quality criteria (e.g., S30).
Genre	Investigates the peer review report as an academic genre, focusing on its overall structure, discourse moves, and the way judgments, comments, and recommendations are organized and communicated.	Studies analyzing structural moves, schematic organization, overall macro-structure, or formulaic lexical bundles.	Studies focusing on the semantic topic of the comments without analyzing the report's overall rhetorical structure.	Identifying typical rhetorical moves (e.g., summarizing, critiquing, recommending) in review reports (e.g., S20, S28).

Theme	Operational Definition	Inclusion Criteria	Exclusion Criteria	Example from Dataset
Metadiscourse	Involves the investigation of the use of different metadiscourse devices, whether interactional or evaluative, to project a social presence and manage interactions with authors.	Studies explicitly utilizing metadiscourse/stance frameworks (e.g., Hyland, 2005) to quantify markers (e.g., hedges, boosters, self-mentions).	Studies analyzing broad speech acts or politeness without a specific focus on defined metadiscourse/stance markers.	Quantifying the frequency of hedges, boosters, and attitude markers across disciplines (e.g., S8).
Association	Explores links between characteristics of reviewer comments and contextual factors, such as editorial decisions, reviewer agreement, or manuscript quality. This may involve statistical correlations or descriptive patterns.	Studies aiming to establish correlations or direct links between feedback patterns and external variables/outcomes (e.g., rejection).	Studies that merely describe comment types without linking them to external variables or subsequent editorial outcomes.	Statistically testing the correlation between review length and the final verdict, or geography and acceptance (e.g., S4, S5).

The following sections present the findings for each theme across the 33 included studies:

Comparative (How Review Comments Vary Across Groups and Conditions):

Out of thirty-three included studies, twenty-four (72.7%) focused on this theme. The main comparative areas were editorial outcomes (e.g., accepted vs. revised vs. rejected) (S1-S3, S14, S21- S24, S26, S27), disciplinary variation (S1, S7, S8, S11, S13, S31, S33), author or reviewer background (e.g., native vs. non-native / Anglophone vs. non-Anglophone / EIL vs. EL) (S2, S3, S5, S10, S25), cultural variation (e.g., British vs. Russian) (S18, S19), and genre modalities (e.g., written peer reviews vs. spoken PhD defenses, or post-publication article comments vs. pre-publication reviews) (S12, S15, S32).

Pragmatics (How Reviewers Do Evaluation, i.e., Politeness, Directives, Persuasion, Roles):

Sixteen studies (48.5%) investigated the pragmatics of reviewers' comments. Focusing on interpersonal dynamics of reviewer comments, specific attention has been paid to politeness/impoliteness and mitigation (S2, S18, S19, S24, S27, S31, S32, S33), speech acts (S7, S21, S31, S32), and persuasive strategies (S15). Moreover, several studies (S9, S17, S22, S29, S31) dealt with the broader stance/role positioning, mapping how reviewers use language to demonstrate their professional identities as gatekeepers versus mentors, facilitators, or advisors. Additionally, some studies explicitly examined evaluative tone, sentiment, and valence (S1, S20).

Comment Focus (Targets of Critique and Language-versus-Content Emphases):

Sixteen studies (48.5%) traced the particular objects of reviewers' criticism (S2-S6, S9, S10, S13, S20, S23, S28-S33). These studies mainly focused on broad content versus language distinctions (S13, S28, S33), as well as specific, language-focused accounts of linguistic problems (S3, S10). Other studies focused on the allocation of comments to particular parts of the manuscript (S2, S4, S5, S6, S20, S23, S29-S32), and the particular research quality standards, such as clarity, originality, and methodological quality, highlighted in the comments (S4, S5, S9, S13, S30-S33).

Genre (Structure of Reviewer Reports):

Fourteen studies (42.4%) considered the peer review report as an occluded academic genre and investigated its macro-structural design and linguistic building blocks. This genre-related research usually focused on move/schematic-level structure and discourse functions (such as summarizing judgments, comments, and recommendations) (S1, S2, S6, S7, S16, S20, S23, S27, S28, S31, S32), and in some studies, on formulaic lexical bundles as phraseological

patterns of reviewer discourse (S11, S17). One study explored the concept of “genre” in a broader sense as the peer-review cycle as a process/practice, not limited to the report structure (S29). Some studies also explored how positive vs. negative evaluations are allocated to different parts of the report (S1, S2, S20, S23, S27, S31, and S32).

Metadiscourse (Stance and Engagement Resources in Reviewer Language):

Six studies (18.2%) employed the metadiscourse and/or stance-engagement perspectives of Hyland (2005) to examine interactional features in reviewer reports. This theme included the quantification of interactional stance features (e.g., hedges, boosters, attitude markers, self-mentions) (S8, S12, S22, S25) and engagement strategies (e.g., reader pronouns, directives, questions, personal asides) (S14, S26) to explore how reviewers position themselves in relation to authors /readers through evaluative discourse.

Association (Correlation/Relationship Patterns):

Five studies (15.2%) directly investigated whether characteristics of reviewers’ comments were related to editorial decisions or contextual factors. Some studies statistically tested these relationships (e.g., S4-S5 tested links with verdict, length of review, agreement of reviewers, geography, and language standard; S33 tested the link between discipline and type of comment). By contrast, other studies made descriptive links (e.g., S30 proposes that rejection is a consequence of the cumulative diversity of criticisms; S3 relates profiles of language-related feedback to acceptance and rejection rates). Figure 5 presents the included studies’ thematic orientations.

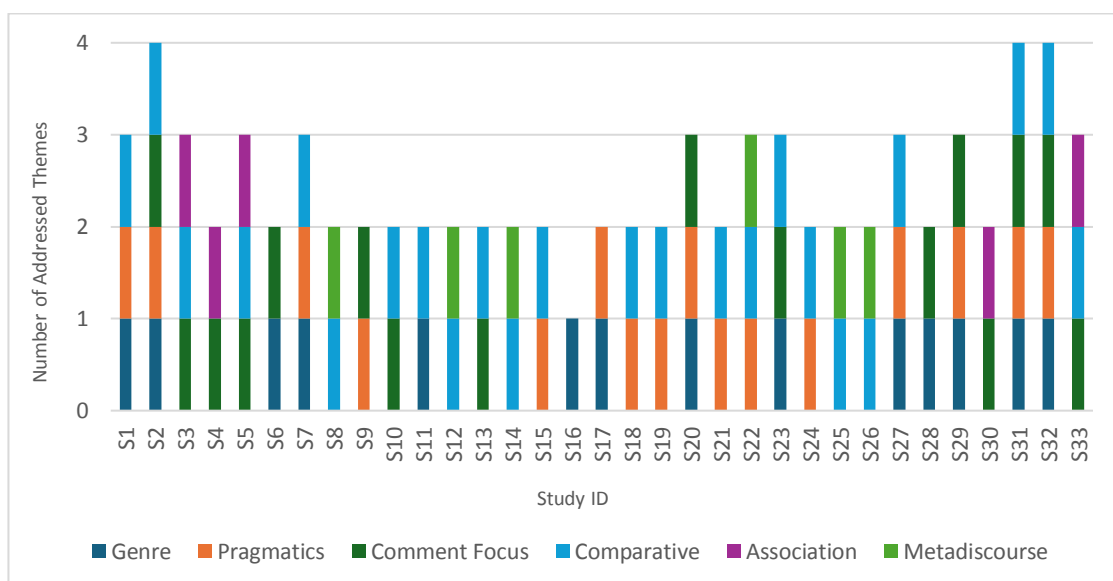


Figure 5. The Included Studies’ Thematic Orientations

As is evident in the above chart, most studies showed a multi-thematic approach. The overall frequency of each theme is also presented in Figure 6.

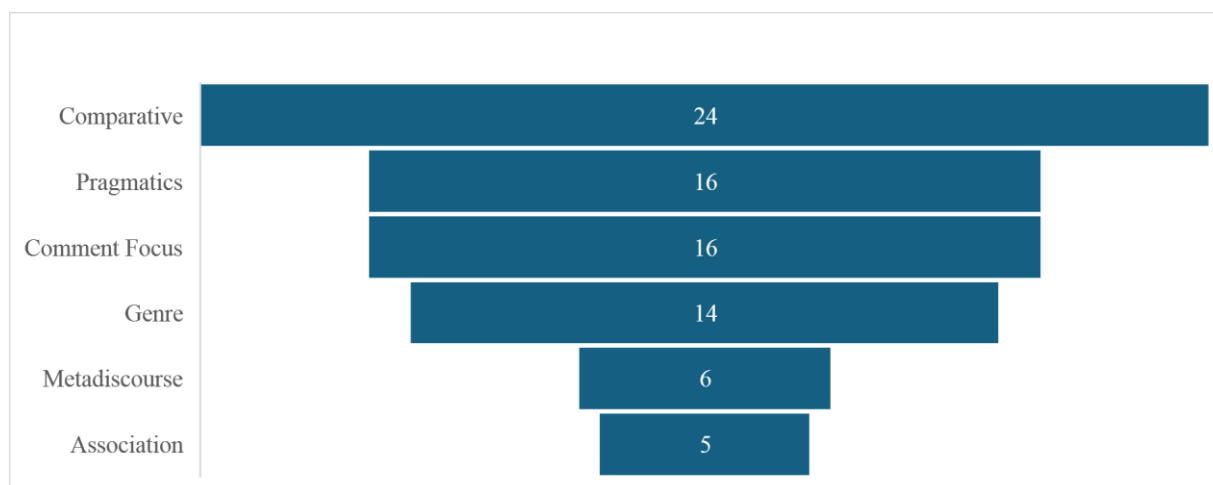


Figure 6. Total Frequency of Individual Themes Addressed in the Studies

Research Question Three: Methodological Characteristics of Included Studies Including Paradigms, Designs, Data Collection, and Analytic Techniques

The third research question sought to identify the research paradigms, research designs, data collection methods and instruments, and analytical techniques employed in the included primary studies. As regards research paradigms, the qualitative approach predominated (27/33) (S1–S3, S6–S8, S11, S13–S32). The mixed-methods approach was the next most common, appearing in four studies ($n = 4$; S4, S5, S12, S33). A purely quantitative approach was only utilized by two studies ($n = 2$; S9, S10). Figure 7 provides an overview of the paradigms.

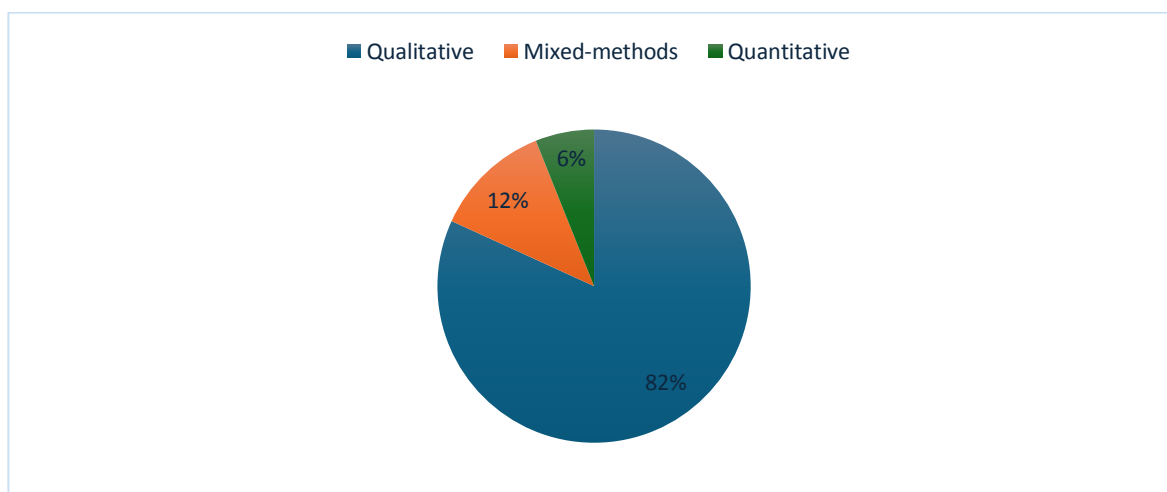


Figure 7. The Distribution of Paradigms in the Included Studies

Within the three paradigms, the studies employed a variety of specific research designs. Descriptive comparative design was the most dominant one ($n = 21$; 63.6%), used to compare variables such as editorial outcomes, disciplines, or author backgrounds (S1–S3, S7, S8, S10, S11, S13–S15, S18, S19, S21–S27, S31, S32). The next top design was descriptive, with four studies ($n = 4$; 12.1%) utilizing a purely descriptive design (S6, S9, S16, S20). Four qualitative studies ($n = 4$; 12.1%) were categorized as case studies, often focusing deeply on a single journal or a specific author's own review history (S17, S28, S29, S30). Furthermore, three mixed-methods studies ($n = 3$; 9.1%) used a convergent design to analyze qualitative and quantitative data concurrently (S4, S5, S12). One study ($n = 1$; 3.0%) also utilized an explanatory sequential design, following up document collection with questionnaires and interviews (S33). Table 6 displays the details of the research paradigms and designs of the included studies.

Table 6. Research Paradigms and Designs Utilized in the Included Studies

Category	Frequency (n)	Percentage (%)	Study IDs
Research Paradigm			
Qualitative	27	81.8%	S1–S3, S6–S8, S11, S13–S32
Mixed-methods	4	12.1%	S4, S5, S12, S33
Quantitative	2	6.1%	S9, S10
Research Design			
Descriptive Comparative	21	63.6%	S1–S3, S7, S8, S10, S11, S13–S15, S18, S19, S21–S27, S31, S32
Descriptive	4	12.1%	S6, S9, S16, S20
Case Study	4	12.1%	S17, S28, S29, S30
Convergent	3	9.1%	S4, S5, S12
Explanatory sequential	1	3.0%	S33

Regarding the data collection methods and instruments, all studies (33/33; 100%) relied on document collection of peer review reports/texts as a core data source. In addition, five studies (5/33; 15.2%) incorporated supplementary data sources/instruments beyond peer review reports. These included transcripts of oral examinations/PhD defenses (2/33; 6.1%) (S12, S15), questionnaires (2/33; 6.1%) (S21: open-ended questionnaire, S33: online questionnaire), Interviews (1/33; 3.0%) (S33: semi-structured interviews), and post-

publication article comments as an additional text type (1/33; 3.0%) (S32). These details are illustrated in Table 7.

Table 7. Data Sources and Collection Instruments

Data Source / Instrument	Frequency (n)	Percentage (%)	Study IDs
Document collection (Peer review reports)	33	100.0%	S1–S33
Transcripts of oral examinations / PhD defenses	2	6.1%	S12, S15
Questionnaires	2	6.1%	S21, S33
Semi-structured interviews	1	3.0%	S33
Article comments (Post-publication)	1	3.0%	S32

Concerning the corpus size (peer review reports), based on the extracted sample sizes for peer review reports, corpora ranged from 18 to 268 reports in most studies (median 82 reports), with one study reporting a macro-corpus of >1,000 peer review reports. These sample sizes are shown in Figure 8.

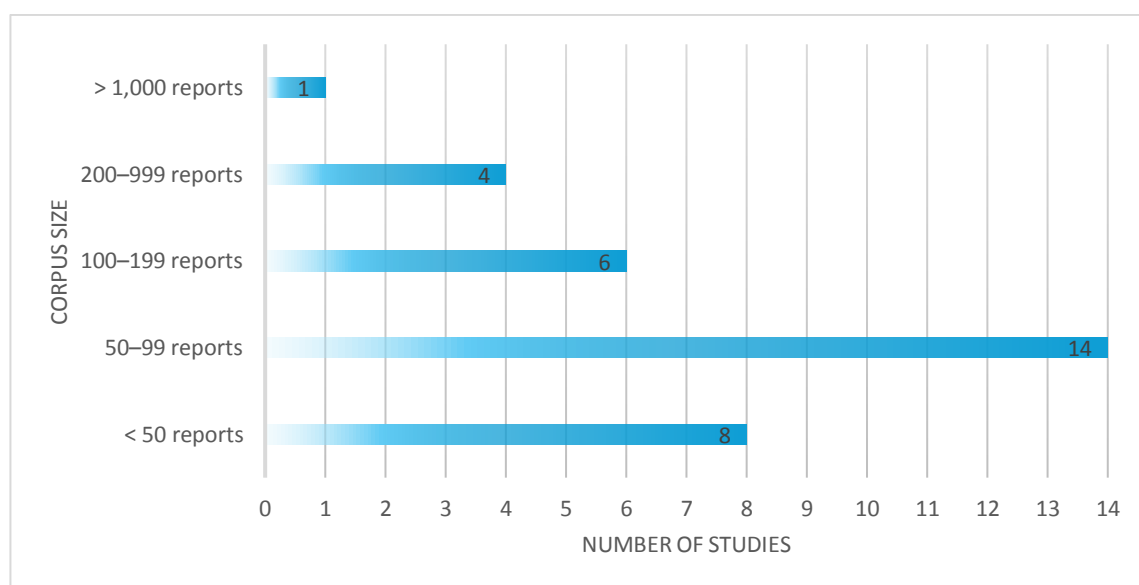


Figure 8. The Corpus Size of the Included Studies

The researchers resorted to multiple analytical methods to study the collected data. Accordingly, a single study may contain more than one analytical method. Discourse Analysis was the most prevalent qualitative technique, utilized by 24 studies (72.7%) to examine language use, moves, and interpersonal dynamics (S1, S2, S6–S8, S11–S19, S21–S28, S31, S32). Twelve studies (36.4%) utilized content analysis to categorize the

targets of reviewer critiques and identify themes (S1–S5, S13, S20, S23, S28–S30, S33). Grounded theory, thematic analysis, and narrative analysis were occasionally incorporated into these broader qualitative coding frameworks in studies S4, S33, and S29, respectively. The descriptive statistics (such as frequency counts and percentages) were highly prevalent, appearing in 30 out of 33 studies (90.9%). Nevertheless, their role varied based on the research design. In the vast majority of studies ($n = 28$), descriptive statistics were used only as a secondary and supportive technique to quantify the occurrence of linguistic features or comment types that were derived from the primary qualitative analyses. However, in the two purely quantitative studies (S9, S10), descriptive statistics served as the main analytic method to map the distribution of reviewer feedback. The details are provided in Table 8.

Table 8. Data Analytic Methods Utilized in the Included Studies

Analytic Method	Frequency (n)	Percentage (%)	Study IDs
Discourse analysis	24	72.7%	S1, S2, S6–S8, S11–S19, S21–S28, S31, S32
Content analysis	12	36.4%	S1–S5, S13, S20, S23, S28–S30, S33
Descriptive statistics (Total)	30	90.9%	
<i>As primary analytic method</i>	2	6.1%	S9, S10
<i>As companion/supplementary method</i>	28	84.8%	S1–S8, S11–S15, S17, S20–S33
Inferential statistics	4	12.1%	S4, S5, S12, S33

Finally, a minority of studies ($n = 4$; 12.1%) ran inferential statistical tests. These tests, which are presented in Table 9, were mainly done to establish significant associations or group differences.

Table 9. Inferential Statistical Tests Used in the Studies

Statistical tests/procedures reported	Study IDs
Pearson correlation	S4
Cohen's kappa	S4, S12
Kruskal–Wallis one-way ANOVA	S4, S5
Post-hoc Scheffé tests	S4
Chi-square test of association	S5, S33
Log-likelihood (LL) ratio tests	S12

Discussion

The aim of this study was to map and synthesize the literature on peer review reports in applied linguistics. The researchers worked on 33 empirical studies (2004-2025) to explore their publication volume and dispersion, dominant themes, and methodological traits.

Extent and Distribution of the Literature (RQ1)

One of the major findings was that the number of studies focusing merely on blind peer review reports in applied linguistics is still relatively small (as presented in the PRISMA flow diagram) but has been expanding noticeably in recent years. Although the studies included in this review range from 2004 to 2025, only a few studies were published before 2010.

The number of publications increased significantly after 2015, with a clear peak in 2020-2025. This increase mirrors broader concerns regarding the viability and productivity of the peer review process in an academic context that encourages researchers to publish more (Drozd & Lodomery, 2024; Publons, 2018).

Most of the studies were published in the form of journal articles (69.7%), while the remaining were published as book chapters in edited volumes and monographs (30.3%). This might be due to the publish-or-perish reward system, where journal articles are viewed as more legible and valued in the review, promotion, and tenure processes (Niles et al., 2020). Consistent with this explanation, the number of journal articles per person grew from 2011 to 2019, while the number of books per person dropped significantly across all disciplines (Savage & Olejniczak, 2022).

In addition, studies are being published at various venues and not being centralized to specific multidisciplinary journals, with 27 different publication venues in 33 studies. On the one hand, this variety shows the interdisciplinary appeal, as scholars from the fields of academic writing, publication pedagogy, and research evaluation are drawn to this area. On the other hand, it shows the methodological diversity of this area, as different venues focus on different lenses (e.g., genre analysis, pragmatics, or corpus linguistics). This finding corroborates the observation that peer review practices are often highly idiosyncratic and dependent on local disciplinary communities (Bornmann, 2011; Bornmann & Mungra, 2011).

More importantly, in terms of data sources, nearly half of the studies (48.5%) obtained their data from a single journal, with *English for Specific Purposes (ESPJ)* being a prominent source of data (30.3%). This is hardly surprising, given EAP's historical focus on occluded academic genres and publication practices. Such reliance on a single journal, however, comes with generalizability issues. This, however, is in contrast with the much larger corpus sizes

investigated in other disciplines, such as [Garcia-Costa et al. \(2022\)](#), who analyzed 1.3 million first-round review reports from 740 Elsevier journals, and [Buljan et al. \(2020\)](#), who analyzed 472,449 peer review reports across 61 journals. [Bornmann et al. \(2010\)](#) showed that peer review is not a highly reliable process and that different referees often make incongruent decisions about the same paper. Thus, findings that are largely dependent on a single journal would be most usefully interpreted as reflecting norms of the journal rather than those of reviewers *per se*.

Moreover, while 24.2% of studies utilized multi-journal datasets that were clearly specified, 27.3% of studies utilized unspecified sources. Therefore, research relying on undisclosed venues creates a methodological blind spot since researchers cannot synthesize the results or replicate the analysis without knowing where the data came from. It also indicates that peer review reports are part of an occluded genre and are not disclosed in the public record by nature (Swales, 1996, as cited in [Gosden, 2003](#)).

In addition, a very strong English-dominant preference was discovered, with 31 of 33 studies conducted only on English manuscripts' review reports and comments. To multilingual academics, the communicative burden of interpreting and addressing the comments is proportionally high ([Mudrak, 2013](#)). However, this focus on English is thoroughly justifiable, as most journals are published in English, and it is the main lingua franca of the world.

Thematic Focus (RQ2)

The findings reflected the portrayal of peer review as a socially constructed interpersonal phenomenon. The themes Comparative (72.7%) and Pragmatics (48.5%) dominated the literature. Therefore, it is evident that the papers focused heavily on the ways feedback differs between groups (editorial decisions, disciplines, author demographics, etc.), and how reviewers navigate their complex dual roles. Accordingly, this approves of the argument by [Tumin and Tobias \(2019\)](#) and [Kelly et al. \(2014\)](#), who suggest that peer reviewers must walk a tightrope between establishing themselves as gatekeepers and supporters.

Another clear pattern of thematic co-occurrence supporting this social orientation is the fact that the studies whose thematic focus was metadiscourse (18.2%) were always coupled with comparative themes. This suggests that interactional features (e.g., hedges, boosters, and attitude markers) are hardly ever the object of a separate study. On the contrary, they are constantly utilized in comparison with different editorial outcomes or NNES/NES author backgrounds in order to uncover reviewer biases in the deployment of authorial stance.

Apart from these interpersonal relationships, a considerable extent of studies was devoted to Comment Focus (48.5%) and Genre (42.4%), which indicates that EAP and genre strongly guide researchers in their efforts to discover these hidden texts. More specifically, researchers were particularly interested in the difference between language and content comments in the Comment Focus category. In other words, they sought to address whether reviewers were acting as proofreaders concerned with superficial grammar and style, or as gatekeepers with respect to deeper issues, such as methodological and theoretical contributions of the papers. Essentially, this category of study is an answer to [Chong and Mason's \(2021\)](#) call to demystify hidden academic genres.

Furthermore, researchers frequently combined Genre, Comment Focus, and Pragmatics within the same studies. By triangulating these three themes, scholars treated the peer review report as a complete, interconnected textual ecosystem. In doing so, they hoped to grasp not only the focus of the critique (content versus language) and its rhetorical positioning (move structure), but also to determine the ways in which it is softened through interpersonal strategies like hedging or indirectness.

However, an important research gap was discovered through thematic mapping. Despite the broad emphasis on the structure and object of the reviewers' comments, only a few studies (15.2%) examined Association themes. This is in marked contrast to the recent large-scale scientometric studies in the medical and hard sciences (e.g., [Buljan et al., 2020](#); [Garcia-Costa et al., 2022](#); [Falk Delgado et al., 2019](#)), which often relied on large data sets to directly associate the reviewers' language with the final editorial outcome. The lack of emphasis on association aspects is consistent with the results of the scoping review of language testing by [He et al. \(2025\)](#). Without further analysis of association patterns, it would be difficult to determine the effect of particular patterns of feedback or mitigation approaches on manuscript outcomes.

Methodological Traits (RQ3)

As regards the research approach and design, there was a clear dominance of the qualitative paradigm (81.85%), with discourse analysis (72.7%) being the primary analytical method and nearly two-thirds of the studies employing a descriptive comparative design (63.6%). These methodological preferences align with the findings of research question two. Since researchers in this area are interested in the rhetorical organization of peer review reports along with the interpersonal aspects of evaluative feedback, qualitative discourse-analytic methods are the best fit for investigating these issues in their contexts. This is in line with the

notion that the use of qualitative approaches is on the rise in applied linguistics, specifically as it gives a more nuanced understanding of the phenomenon of interest (Alisoy, 2025).

The heavy use of descriptive comparative designs in the majority of the studies may reflect the field's aim to examine whether there are differences in reviewers' comments based on their nativeness and the final publication outcome. Within the predominantly qualitative approaches to analyzing the research articles, however, there was a split in the methods researchers used to process the texts. While discourse analysis was chosen as the main method for analyzing the texts' microlinguistic elements (e.g., speech acts, hedges, moves), content analysis was also used in 36.4% of the studies. These two methods indicate that peer review is conceptualized in two main ways within applied linguistics: how the interaction is expressed (discourse) and what the critique actually addresses (content).

The sizes of the corpora utilized in the selected studies, however, were relatively small. This is evident in that the majority of studies examined fewer than 100 research reports (66.6%), with only one study utilizing over 1,000 research reports within its macro-corpus. This pattern of using smaller review corpora reflects both the labor involved in performing qualitative analysis on research articles and the difficulty in gaining access to blind peer review reports in large quantities. Consequently, quantitative research, which generally necessitates larger corpora, is rare within the field altogether.

Finally, there was a limited use of advanced statistical testing across the included studies. Despite the high prevalence of descriptive statistics (90.9%), they were used almost exclusively as a supplementary technique in the forms of frequency and percentages to summarize the results of the qualitative analysis. Only a minor number of studies (12.1%) employed inferential statistical tests (e.g., Kruskal–Wallis, Chi-square, or Cohen's kappa) to test the significant associations or group differences.

This relatively limited incorporation of advanced statistical analysis should not be considered simply a matter of disciplinary preference, but rather a direct function of the type of data. Due to the small, naturalistic data sets that are often used, these studies may not have the power or the standardized variable types required for advanced quantitative analyses. This heavy dependence on descriptive statistics may limit the field's ability to move beyond exploratory mapping into predictive modeling.

Conclusion

This scoping review sought to provide an overview of the current landscape of literature on peer review comments published in the field of applied linguistics in terms of extent, themes,

and methodologies. We discovered that although the field is still rather narrow, it has been expanding in recent years. This increase suggests a growing appreciation among academics for the role of review in academic work. The 33 included studies were dispersed across many publication outlets, showed a strong preference for English-language reports, and often relied on a limited set of reports as their corpus.

From the thematic mapping, we identified four primary areas of focus in studies on peer review comments, namely comparisons, pragmatics, comment focus, and genre. The reviewed studies also treated the peer review report as a socially constructed, evaluative, and systematic text rather than as a tool for objective measurement of scientific quality. Regarding methodology, we found qualitative research paradigms, discourse analysis, and descriptive comparative designs to be highly prevalent.

Moreover, our review revealed a lack of research connecting reviewer language to editorial outcomes. This may point to a significant lack of insight into how particular feedback features affect a manuscript's final acceptance or rejection.

Furthermore, the exclusive use of an English-language corpus, combined with limited sample sizes and inadequate reporting of data collection periods, makes it difficult to generalize the findings and compare them across publication contexts and time periods. Expanding the scope to include larger, multilingual datasets from different journals, along with mixed-methods approaches or statistical analyses where applicable, would significantly enrich this area of research.

Overall, to our knowledge, this review provides the first consolidated overview of how peer review comments have been studied in applied linguistics, offering a foundation for future work in this area. However, this scoping review suffers from some limitations. Firstly, although we searched major databases (Scopus and Web of Science) and conducted hand searches, some relevant studies may have been omitted due to indexing limitations or lack of access to full texts. Secondly, due to the nature of scoping reviews, we did not seek to assess the methodological quality of the included studies or synthesize effect sizes. Thirdly, this study focused on blind review comments and excluded open review reports to maintain a focused scope. Finally, our inclusion of only English-language publications means that studies published in other languages, including those analyzing non-English peer review practices, may have been missed.

Future secondary research can benefit from mapping other aspects, such as reviewers' and writers' opinions on the publication process, writers' responses to reviewers, or even

open peer review comments. They could also map the findings of the literature to find the recurrent patterns, areas of convergence and divergence, and broader trends across studies.

Acknowledgement

The authors would like to express their gratitude to all those who supported this research.

Declaration of Conflicting Interests

The authors declare no conflicts of interest.

Funding Details

This research did not receive any funding.

References

- Aczel, B., Barwich, A. S., Diekman, A. B., Fishbach, A., Goldstone, R. L., Gomez, P., ... & Ioannidis, J. P. (2025). The present and future of peer review: Ideas, interventions, and evidence. *Proceedings of the National Academy of Sciences*, 122(5), e2401232121. <https://doi.org/10.1073/pnas.2401232121>
- Alisoy, H. (2025). Evolving research methodologies in Applied Linguistics: A conceptual analysis. *Porta Universorum*, 1(2), 47-55. <https://doi.org/10.69760/portuni.010206>
- American Journal Experts. (2018). *Peer review: How we found 15 million hours of lost time*. Retrieved December 18, 2025, from <https://www.aje.com/en/arc/peer-review-process-15-million-hours-lost-time>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Atjonen, P. (2019). Peer review in the development of academic articles: Experiences of Finnish authors in the educational sciences. *Learned Publishing*, 32(2), 137–146. <https://doi.org/10.1002/leap.1204>
- Bahl, R., & Bahl, S. (2021). Publication pressure versus ethics, in research and publication. *Indian Journal of Community Medicine*, 46(4), 584–586. https://doi.org/10.4103/ijcm.IJCM_309_20
- Belcher, D. D. (2007). Seeking acceptance in an English-only research world. *Journal of Second Language Writing*, 16(1), 1–22. <https://doi.org/10.1016/j.jslw.2006.12.001>

- Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, 45, 197–245. <https://doi.org/10.1002/aris.2011.1440450112>
- Bornmann, L., & Mungra, P. (2011). Improving peer review in scholarly journals. *European Science Editing*, 37(2), 41–43. <https://europeanscienceediting.org.uk/wp-content/uploads/2017/04/esemay2011.pdf>
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2010). A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants. *PLOS ONE*, 5(12), e14331. <https://doi.org/10.1371/journal.pone.0014331>
- Buljan, I., Garcia-Costa, D., Grimaldo, F., Squazzoni, F., & Marušić, A. (2020). Large-scale language analysis of peer review reports. *eLife*, 9, e53249. <https://doi.org/10.7554/eLife.53249>
- Chong, S. W., & Mason, S. (2021). Demystifying the process of scholarly peer-review: An autoethnographic investigation of feedback literacy of two award-winning peer reviewers. *Humanities and Social Sciences Communications*, 8(1), 1–11. <https://doi.org/10.1057/s41599-021-00951-2>
- Cobey, K. D., Ebrahimzadeh, S., Page, M. J., Thibault, R. T., Nguyen, P.-Y., Abu-Dalfa, F., & Moher, D. (2024). Biomedical researchers' perspectives on the reproducibility of research. *PLOS Biology* 22(11), e3002870. <https://doi.org/10.1371/journal.pbio.3002870>
- Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: The SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research*, 22(10), 1435–1443. <https://doi.org/10.1177/1049732312452938>
- Drozd, J. A., & Ladomery, M. R. (2024). The peer review process: Past, present, and future. *British Journal of Biomedical Science*, 81, 12054. <https://doi.org/10.3389/bjbs.2024.12054>
- Ellinger, A. D., Anderson, V., Gubbins, C., Lynn Lunn, M., Nimon, K. F., Sheehan, M., & Werner, J. M. (2013). The generous spirit of the peer review process: Perspectives and insights from the HRDQ editorial team on providing high quality reviews. *Human Resource Development Quarterly*, 24(4), 417–428. <https://doi.org/10.1002/hrdq.21176>
- Falk Delgado, A., Garretson, G., & Falk Delgado, A. (2019). The language of peer review reports on articles published in the *BMJ*, 2014–2017: An observational study. *Scientometrics*, 120, 1225–1235. <https://doi.org/10.1007/s11192-019-03160-6>
- Fortanet-Gómez, I. (2009). Strategies for teaching and learning an occluded genre: The RA referee report. In S. Burgess & P. Martín-Martín (Eds.), *English as an additional*

- language in research publication and communication* (Linguistic Insights: Studies in Language and Communication, Vol. 61, pp. 19–38). Peter Lang. <https://www.peterlang.com/document/1104303>
- Ganji, M., & Derakhshan, A. (2020). Developing a checklist for evaluating research articles in applied linguistics. *Teaching English Language*, 14(2), 239-268. <https://doi.org/10.22132/tel.2020.121858>
- Garcia-Costa, D., Squazzoni, F., Mehmani, B., & Grimaldo, F. (2022). Measuring the developmental function of peer review: A multi-dimensional, cross-disciplinary analysis of peer review reports from 740 academic journals. *PeerJ*, 10, e13539. <https://doi.org/10.7717/peerj.13539>
- Glonti, K., Cauchi, D., Cobo, E., Boutron, I., Moher, D., & Hren, D. (2017). A scoping review protocol on the roles and tasks of peer reviewers in the manuscript review process in biomedical journals. *BMJ Open*, 7(10), e017468. <https://doi.org/10.1136/bmjopen-2017-017468>
- Gosden, H. (2003). “Why not give us the full story?”: Functions of referees’ comments in peer reviews of scientific research papers. *Journal of English for Academic Purposes*, 2(2), 87–101. [https://doi.org/10.1016/S1475-1585\(02\)00037-1](https://doi.org/10.1016/S1475-1585(02)00037-1)
- He, S., Sénécal, A. M., Stansfield, L., & Suvorov, R. (2025). A scoping review of research on second language test preparation. *Language Testing*, 42(1), 11–47. <https://doi.org/10.1177/02655322241249754>
- Hewings, M. (2004). An 'important contribution' or 'tiresome reading'? A study of evaluation in peer reviews of journal article submissions. *Journal of Applied Linguistics*, 1(3), 247-274. <https://doi.org/10.1558/japl.2004.1.3.247>
- Hillman, S., Selvi, A. F., & Yazan, B. (2021). A scoping review of World Englishes in the Middle East and North Africa. *World Englishes*, 40(2), 159–175. <https://doi.org/10.1111/weng.12505>
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–192. <https://doi.org/10.1177/1461445605050365>
- Kashiha, H. (2024). Critical comments in the disciplines: A comparative look at peer review reports in applied linguistics and engineering. *Text & Talk*, 45(2), 227-250. <https://doi.org/10.1515/text-2023-0055>
- Kelly, J., Sadeghieh, T., & Adeli, K. (2014). Peer Review in scientific publications: Benefits, critiques, & a survival guide. *EJIFCC*, 25(3), 227–243. <https://pubmed.ncbi.nlm.nih.gov/27683470/>

- Kheradparvar, N., & Dreyfus, S. (2025). Treading carefully: A genre analysis of “accept with revision” peer reviews of linguistic journal submissions using the appraisal system. *Journal of World Languages*, 11(1), 262–281. <https://doi.org/10.1515/jwl-2024-0062>
- Lu, X. (2022). What drives Chinese scholars to publish in international journals? Motivations and implications. *Higher Education Research & Development*, 41(6), 1977–1991. <https://doi.org/10.1080/07294360.2021.1971162>
- Marina, T., & Sterligov, I. (2021). Prevalence of potentially predatory publishing in Scopus on the country level. *Scientometrics*, 126(6), 5019–5077. <https://doi.org/10.1007/s11192-021-03899-x>
- Mason, S., & Chong, S. W. (2022). Bringing light to a hidden genre: The peer review report. *Higher Education Research & Development*, 42(3), 664–678. <https://doi.org/10.1080/07294360.2022.2073976>
- Mohammadi, V., Amini Farsani, M., & Nazmi, R. (2023). Peer reviewing in applied linguistics: Reviewers’ perceptions. *Interdisciplinary Studies in English Language Teaching*, 1(1), 53–64. <https://doi.org/10.22080/iselt.2021.21088.1010>
- Mudrak, B. (2013). Understanding the needs of international authors. *Learned Publishing*, 26(2), 139–147. <https://doi.org/10.1087/20130212>
- Niles, M. T., Schimanski, L. A., McKiernan, E. C., & Alperin, J. P. (2020). Why we publish where we do: Faculty publishing values and their relationship to review, promotion and tenure expectations. *PLOS ONE*, 15(3), e0228914. <https://doi.org/10.1371/journal.pone.0228914>
- Paltridge, B. (2015). Referees’ comments on submissions to peer-reviewed journals: When is a suggestion not a suggestion? *Studies in Higher Education*, 40(1), 106–122. <https://doi.org/10.1080/03075079.2013.818641>
- Paltridge, B. (2017). *The discourse of peer review: Reviewing submissions to academic journals*. Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-48736-0>
- Paruzel-Czachura, M., Baran, L., & Spendel, Z. (2021). Publish or be ethical? Publishing pressure and scientific misconduct in research. *Research Ethics*, 17(3), 375–397. <https://doi.org/10.1177/1747016120980562>
- Peters, M. D. J., Godfrey, C., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2020). Scoping reviews. In E. Aromataris, C. Lockwood, K. Porritt, B. Pilla, & Z. Jordan (Eds.), *JBIM Manual for Evidence Synthesis* (2024 ed.). JBI. <https://doi.org/10.46658/JBIMES-24-09>

- Publons. (2018). *Global state of peer review 2018*. <https://publons.com/static/Publons-Global-State-Of-Peer-Review-2018.pdf>
- Ross-Hellauer, T., & Horbach, S. P. J. M. (2024). Additional experiments required: A scoping review of recent evidence on key aspects of Open Peer Review. *Research Evaluation*, 33, rvae004. <https://doi.org/10.1093/reseval/rvae004>
- Ross-Hellauer, T., Bouter, L. M., & Horbach, S. P. J. M. (2023). Open peer review urgently requires evidence: A call to action. *PLOS Biology*, 21(10), e3002255. <https://doi.org/10.1371/journal.pbio.3002255>
- Savage, W. E., & Olejniczak, A. J. (2022). More journal articles and fewer books: Publication practices in the social sciences in the 2010's. *PLOS ONE*, 17(2), e0263410. <https://doi.org/10.1371/journal.pone.0263410>
- Sizo, A., Lino, A., Rocha, Á., & Reis, L. P. (2025). Defining quality in peer review reports: A scoping review. *Knowledge and Information Systems*, 67(8), 6413–6460. <https://doi.org/10.1007/s10115-025-02435-0>
- Song, E., Ang, L., Park, J.-Y., Jun, E.-Y., Kim, K. H., Jun, J., Park, S., & Lee, M. S. (2021). A scoping review on biomedical journal peer review guides for reviewers. *PLOS ONE*, 16(5), e0251440. <https://doi.org/10.1371/journal.pone.0251440>
- Sönmez, D., & Akbas, E. (2023). ‘Great work folks!’: Establishing interpersonal communication in transparent peer reviews of research articles. *Ibérica*, (46), 69–95. <https://doi.org/10.17398/2340-2784.46.69>
- Stupacher, J. (2025). Enhancing peer review skills in higher education: A mixed-methods study on challenges and training needs. *Discover Education*, 4, Article 211. <https://doi.org/10.1007/s44217-025-00663-8>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., ... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>
- Tumin, D., & Tobias, J. D. (2019). The peer review process. *Saudi Journal of Anaesthesia*, 13(Supplement 1), S52–S58. https://doi.org/10.4103/sja.SJA_544_18
- Yakhontova, T. (2019). “The authors have wasted their time...”: Genre features and language of anonymous peer reviews. *Topics in Linguistics*, 20(2), 67–89. <https://doi.org/10.2478/topling-2019-0010>

Appendix

Appendix A

The List of the Included Studies

- S1. Al-Khasawneh, F. (2022). Analysis of the language used in the reports of peer-review journals. *Applied Research on English Language*, 11(3), 79-94. <https://doi.org/10.22108/are.2022.130458.1774>
- S2. Belcher, D. D. (2007). Seeking acceptance in an English-only research world. *Journal of Second Language Writing*, 16(1), 1-22. <https://doi.org/10.1016/j.jslw.2006.12.001>
- S3. Bocanegra-Valle, A. (2015). Peer reviewers' recommendations for language improvement in research writing. In R. Plo Alastrué & C. Pérez-Llantada (Ed.), *Volume 2 English as a Scientific and Research Language: Debates and Discourses* (pp. 207-230). Berlin, München, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9781614516378-012>
- S4. Coniam, D. (2011). Systematising system: One reviewer's analysis of the review process. *System*, 39(4), 539-553. <https://doi.org/10.1016/j.system.2011.10.018>
- S5. Coniam, D. (2012). Exploring reviewer reactions to manuscripts submitted to academic journals. *System*, 40(4), 544-553. <https://doi.org/10.1016/j.system.2012.10.002>
- S6. El Sakran, T., Nunn, R., & Adamson, J. (2019). A genre analysis of the schematic structure and linguistic features of reviewers' reports on research manuscripts. *Asian ESP Journal*, 15(3), 7-55. https://www.researchgate.net/publication/337973405_A_Genre_Analysis_of_the_Schematic_Structure_and_Linguistic_Features_of_Reviewers'_Reports_on_Research_Manuscripts
- S7. Fortanet, I. (2008). Evaluative language in peer review referee reports. *Journal of English for Academic Purposes*, 7(1), 27-37. <https://doi.org/10.1016/j.jeap.2008.02.004>
- S8. Fortanet-Gomez, I., & Ruiz-Garrido, M. F. (2010). Interacting with the research article author: Metadiscourse in referee reports. In *Constructing interpersonality: Multiple perspectives on written academic genres* (pp. 243). Springer. https://doi.org/10.1007/978-3-531-93212-9_12
- S9. Hewings, M. (2004). An 'important contribution' or 'tiresome reading'? A study of evaluation in peer reviews of journal article submissions. *Journal of Applied*

- Linguistics*, 1(3), 247-274. <https://doi.org/10.1558/japl.2004.1.3.247>
- S10. Hewings, M. (2006). English language standards in academic articles: Attitudes of peer reviewers. *Revista Canaria de Estudios Ingleses*, 53, 47-62. <http://riull.ull.es/xmlui/handle/915/17190>
- S11. Kashiha, H. (2023). Beyond words in evaluation: Formulaic language in critical reviews of research articles across disciplines. *Russian Journal of Linguistics*, 27(2), 251-275. <https://doi.org/10.22363/2687-0088-34320>
- S12. Kashiha, H. (2024). Stance-taking in peer reviewer and thesis examiner feedback on Iranian scholarly contributions. *Journal of English for Academic Purposes*, 68, 101364. <https://doi.org/10.1016/j.jeap.2024.101364>
- S13. Kashiha, H. (2024). Critical comments in the disciplines: A comparative look at peer review reports in applied linguistics and engineering. *Text & Talk*, 45(2), 227-250. <https://doi.org/10.1515/text-2023-0055>
- S14. Kashiha, H. (2025). From assessment to interaction: Exploring author engagement strategies in manuscript reviews. *Ibérica*, (50), 81–108. <https://doi.org/10.17398/2340-2784.50.81>
- S15. Kashiha, H. (2025). A contrastive analysis of persuasive feedback in written and spoken academic evaluations. *Contrastive Pragmatics*, 6(3), 584-607. <https://doi.org/10.1163/26660393-bja10138>
- S16. Kheradparvar, N., & Dreyfus, S. (2025). Treading carefully: A genre analysis of “accept with revision” peer reviews of linguistic journal submissions using the appraisal system. *Journal of World Languages*, 11(1), 262-281. <https://doi.org/10.1515/jwl-2024-0062>
- S17. Kuteeva, M. (2022). Polycentric peer reviewing: Navigating authority and expertise. In P. Habibie & A.K. Hultgren (Eds.), *The inner world of gatekeeping in scholarly publication* (pp. 101-121). Springer. https://doi.org/10.1007/978-3-031-06519-4_7
- S18. Larina, T., & Ponton, D. (2020). Tact or frankness in English and Russian blind peer reviews. *Intercultural Pragmatics*, 17(4), 471-496. <https://doi.org/10.1515/ip-2020-4004>
- S19. Larina, T., & Ponton, D. (2022). I wanted to honour your journal, and you spat in my face: Emotive (im)politeness and face in the English and Russian blind peer review. *Journal of Politeness Research*, 18(1), 201-226. <https://doi.org/>

10.1515/pr-2019-0035

- S20. Mason, S., & Chong, S. W. (2022). Bringing light to a hidden genre: The peer review report. *Higher Education Research & Development*, 42(3), 664-678. <https://doi.org/10.1080/07294360.2022.2073976>
- S21. Paltridge, B. (2013). Referees' comments on submissions to peer-reviewed journals: When is a suggestion not a suggestion? *Studies in Higher Education*, 40(1), 106–122. <https://doi.org/10.1080/03075079.2013.818641>
- S22. Paltridge, B. (2015). Reviewers' comments on submissions to peer-reviewed journals. In *The discourse of peer review: Reviewing submissions to academic journals* (pp. 115-144). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-48736-0_5
- S23. Paltridge, B. (2017). The genre of reviewers' reports. In *The discourse of peer review: Reviewing submissions to academic journals* (pp. 31-65). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-48736-0_2
- S24. Paltridge, B. (2017). Politeness and reviewers' reports. In *The discourse of peer review: Reviewing submissions to academic journals* (pp. 91-113). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-48736-0_4
- S25. Paltridge, B. (2019). Reviewers' feedback on second-language writers' submissions to academic journals. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 226–244). Cambridge University Press. <https://www.cambridge.org/core/books/abs/feedback-in-second-language-writing/reviewers-feedback-on-secondlanguage-writers-submissions-to-academic-journals/3B9592758BB0134C5A24BBE8645C1F0C>
- S26. Paltridge, B. (2020). Engagement and reviewers' reports on submissions to academic journals. *Journal of English for Research Publication Purposes*, 1(1), 4-27. <https://doi.org/10.1075/jerpp.19007.pal>
- S27. Samraj, B. (2016). Discourse structure and variation in manuscript reviews: Implications for genre categorization. *English for Specific Purposes*, 42, 76-88. <https://doi.org/10.1016/j.esp.2015.12.003>
- S28. Shi, L. (2022). Rhetorical structure and types of comments in my manuscript reviews. In P. Habibie & A.K. Hultgren (Eds.), *The inner world of gatekeeping in scholarly publication* (pp. 101-121). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-06519-4_9
- S29. Tschichold, C., Boulton, A., & Pérez-Paredes, P. (2024). Interpreting the review

- process in applied linguistics research. In *Developing feedback literacy for academic journal peer review: Narratives from researchers in education and applied linguistics* (pp. 173). <https://doi.org/10.4324/9781003365662-5>
- S30. Watson Todd, R. (2013). Why do articles get rejected by international journals? *rEFLections*, 16, 46–57. <https://doi.org/10.61508/refl.v16i0.114212>
- S31. Yakhontova, T. (2019). “The authors have wasted their time...”: Genre features and language of anonymous peer reviews. *Topics in Linguistics*, 20(2), 67–89. <https://doi.org/10.2478/topling-2019-0010>
- S32. Yakhontova, T. (2023). Evaluative genres of research communication: Article comments and peer reviews from linguistic and pedagogical perspectives. In M. Walková (Ed.), *Linguistic approaches in English for academic purposes: Expanding the discourse* (pp. 168–192). Bloomsbury Academic. <https://doi.org/10.5040/9781350300330.0019>
- S33. Zare, J., Mahmoudi-Gahrouei, V., Ketabi, S., & Keivanloo-Shahrestanaki, Z. (2016). English for research publication purposes: The case of scholarly peer review comments. *Ibérica*, 32, 153-178. <http://web.a.ebscohost.com>

Appendix B

Search String for Each Database

Database	Search String
Scopus	TITLE-ABS-KEY ("applied linguistics" OR "linguistics journal*" OR "language education" OR TESOL OR ELT OR EFL OR ESL) AND TITLE-ABS-KEY ("peer review" OR "peer-review" OR "reviewer report*" OR "review report*" OR "reviewer comment*" OR "reviewer feedback" OR "editorial decision letter*") AND TITLE-ABS-KEY (comment* OR feedback OR evaluation OR criticism OR stance OR tone OR directive* OR praise OR "reviewer bias" OR "reviewer impact") AND TITLE-ABS-KEY ("discourse analysis" OR "qualitative study" OR "content analysis" OR "mixed methods" OR "corpus analysis")
Web of Science	Topic ("applied linguistics" OR "linguistics journal*" OR "language education" OR TESOL OR ELT OR EFL OR ESL) AND Topic ("peer review" OR "peer-review" OR "referee report*" OR "review report*" OR "reviewer report*" OR "reviewer comment*" OR "reviewer feedback" OR "editorial decision letter*") AND Topic (comment* OR feedback OR evaluation OR criticism OR stance OR tone OR politeness OR appraisal OR directive* OR hedge*) AND Topic ("discourse analysis" OR "qualitative research" OR "corpus-based analysis" OR "content analysis" OR "mixed methods")

Appendix C

Data Extraction Form

Categories	Variables
Bibliographic and study profile (RQ1)	Study ID
	Publication type
	Journal / Source
	Publication Year
	Author(s)
	Author Affiliation Country
	Target Language
Conceptual and contextual characteristics (RQ2)	Data Collection Period
	Study Context (Country/Region)
Thematic focus and analytical variables (RQ3)	Analytical / Theoretical Framework
	Primary Thematic Focus (allows multiple codes per study)
	+1
	Specific Analytical Framework applied
	Key Variables / Categories Examined