

Students' voice in L2 English writing: A systematic review of literature

Abdelhamid M. Ahmed^{*}, Xiao Zhang

Core Curriculum Program, Deanship of General Studies, Qatar University, Doha, Qatar

ARTICLE INFO

Keywords:

Voice
Metadiscourse
Systematic Review
L2 English
Students' Writing

ABSTRACT

The present systematic review assessed the existing and growing literature in metadiscourse and L2 English writing by focusing on using voice in students' writing. Empirical studies in this field were retrieved and coded for their research design, study types, data sources, software tools, corpora characteristics, study context, analytical tests and reporting practices. Prominent adoption of mixed methods and triangulation approaches was found. Hyland's metadiscourse model (2005) was utilised by most reviewed studies, while models proposed by other researchers were also found in this review, which provide more feasible choices for future studies. A shortage of longitudinal studies of voice and L2 English writing was found, and more evidence from diachronic corpora was spotlighted. Most reviewed studies were corpus-based; however, some studies showed that reporting corpora features remained insufficient. Implications and suggestions for further research of voice markers and L2 English writing were provided based on the review.

1. Introduction

The relationship between voice markers and the quality of L2 English writing is explored in empirical studies inspired by the increasing research on L2 voice markers. There has been increasing research on L2 voice markers and the use of corpora to determine the linguistic and rhetorical markers L2 learners might incorporate into their writing (Ahmed et al., 2023; Myhill et al., 2023; Appel and Szeib, 2018; Granger, 2015; Gong and Barlow, 2022; Hinkel, 2002; Hyland, 2001). This research inspired empirical research on the relationship between using voice markers and the quality of L2 English writing.

The body of literature suggests that voice markers are genre-dependent and context-specific. A body of literature, particularly from Hyland (2000, 2001 & 2002), suggests that voice and authorial identity are genre-dependent. It further underscores a potential need to consider both the intensity and appropriateness of voice within a specific context. While this may be the case, the literature suggests identifying certain linguistic markers that supposedly lend individualised voice to writing or, at the very least, distinguish the writing of L2 English learners from the writing of their L1 counterparts (e.g., Bychkovska & J. Lee, 2017; Hinkel, 1999 & 2002; Wei et al., 2020).

Authors' sense of belonging in writing is expressed through ideology or marker use, while readers infer the intended meaning to interpret the text. This sense of belonging, in some cases, is embodied in authors' willingness to follow certain ideologies, which is often connected with

announcing authors' subjectivity. In other cases, it is reflected in the writers' effort to shape their logic and draws attention to their awareness of using such markers. On the other hand, readers who have no access to the writer's intended meaning can understand and interpret the text coherence and the authors' intention depending on inferences of what the writer means (Flowerdew, 2018).

Voice refers to writer visibility, authorial presence, and stance, but its measurement in L2 English writing is still problematic. In previous research, voice refers to how writers express themselves and their ideas using different terminologies. For instance, the concept is defined as writer visibility, manifested through self-mention, emphatics, evaluative modifiers, and references to the situation of writing (Petch-Tyson, 1998). Another term for voice, authorial presence, shares the same meaning in studies (e.g., Çandarlı et al., 2015; Clark and Ivanić, 1997b; Ivanić and Camps, 2001; Sun et al., 2022). Furthermore, the notion of voice also relates to stance, defined as "the amalgamative effect of the use of discursive and non-discursive features that language users choose, deliberately or otherwise, from socially available yet ever-changing repertoires" (Matsuda, 2001, p. 40). Although the conception and employment of voice are practically the same in L1 and L2 English writing, the voice itself is difficult to measure because "what is measured is limited by what can be measured and by how 'good writing' is conceived in the first place" (Matsuda and Jeffery, 2012, p. 162). Accordingly, linguistic markers can convey an individual's voice, establish the author's text ownership, and are important for second

^{*} Corresponding author.

E-mail addresses: aha202@qu.edu.qa (A.M. Ahmed), zhang.xiao@qu.edu.qa (X. Zhang).

language (L2) student writing. However, whether this strategy improves the quality of a relatively inexperienced L2 writer's texts remains questionable. Thus, to assess the contribution of voice to L2 English students' writing, the present study systematically reviews previous research in this field. It aims to review previous pertinent research in terms of research design, study types, conceptual framework, taxonomies, corpus characteristics and elicitation measures, analytical methods and reporting practices. This study seeks to provide answers and insights to the following research questions:

1. What metadiscourse theoretical frameworks were adopted in the reviewed empirical studies of voice in L2 English writing?
2. What are the methodological characteristics of the reviewed empirical studies of voice in L2 English writing?

2. Method

2.1. Eligibility criteria

Some theoretical frameworks of metadiscourse were proposed (Vande Kopple, 1985; Crismore et al., 1993; Mauranen, 1993; Dahl, 2004; Hyland, 2005a, 2005d, 2005b; Ädel, 2006, 2006d, 2010; Adel and Mauranen, 2010). These frameworks usually led to divergences among metadiscourse studies in L2 English writing in the past decades. The present study adopted a comprehensive and accurate set of search terms selected after considerable discussion with the co-authors. The inclusion/exclusion criteria were built upon a thorough review of theoretical frameworks of metadiscourse and L2 writing and ancestral searching to search through the existing studies.

2.1.1. Inclusion/exclusion criteria

Seven inclusion criteria were followed. Firstly, we reviewed published research from 1985 to 2022. Secondly, accurate and comprehensive search terms were developed using [metadiscourse, voice and writing in the title, abstract and keywords] (see appendix 1 for full search terms). Thirdly, we searched the following databases: WoS, Scopus, EBSCOhost, ERIC, and Google Scholar. Fourthly, we limited our search in Google Scholar to the first 200 hits for manageability purposes. Sixthly, we retrieved documents published as journal articles, conference proceedings, or book chapters and excluded any other wrong document types. Finally, our search lasted three months (August–September 2022).

Eight exclusion criteria were adopted in our search. Firstly, the search hits, either book reviews or position papers, were excluded. Secondly, studies investigating student writing in languages other than L2 English were removed. Thirdly, studies that did not specify participants' language backgrounds were excluded. Fourthly, empirical studies that researched writing by non-native English researchers and ESL adults and the ones which addressed the writing of textbooks, or students' spoken presentations, were excluded. Fifthly, empirical studies that did not consider the metadiscoursal properties were excluded. Sixthly, studies that used voice to address their primary concerns, such as writing assessment, human rating, feedback, or teacher instruction, were removed. Seventhly, studies whose data did not distinguish between writing by L2 English researchers and L2 English students in the analysis were excluded. Finally, studies that mixed data collected from native and non-native English speakers without explanation were excluded.

The researchers followed PRISMA guidelines (Page et al., 2021) in this systematic review. The most important and decisive procedure was to develop an accurate set of search terms to avoid irrelevant studies or fuzzy search results. This was followed by identifying relevant studies in the databases. The process began with a trial retrieval of 32 documents (Articles/Proceedings/Early Access) written in English from Web of Science (WoS) with "voice", "metadiscourse", and "writing" in the title, abstract and keywords. After analysing Author Keywords and Keywords

Plus provided by WoS, further and detailed amendments were made to ensure comprehensive retrieval. The final set of search terms, including contextual keywords, was provided in Appendix (1).

2.2. Information sources and search

The search terms were applied to the chosen research indices' title, abstract and keyword fields in the databases: WoS, Scopus, EBSCOhost, and ERIC. An additional search of Google Scholar was conducted to avoid omitting studies published in journals not indexed in the above databases. We limited the Google Scholar search to the first 200 hits for manageability purposes. Our search reviewed published research for 37 years (1985–2022). 1985 marked the publication of the first metadiscourse model by Vande Kopple. Additionally, September 2022 was the cut-off date, as our search was conducted from August to September 2022.

2.3. Study selection and data extraction process

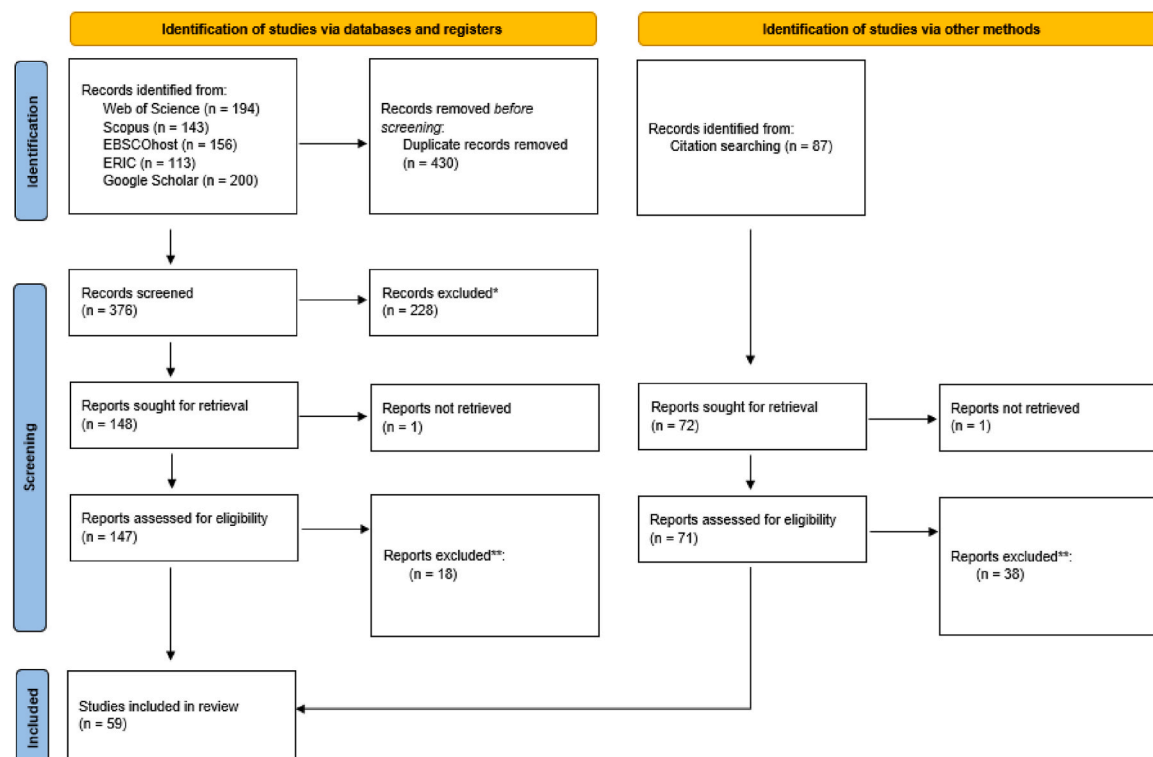
Following PRISMA procedures, the titles and abstracts were screened to determine the relevance of the study after completing the search process (See Fig. 1). The number of duplicate records (i.e., the number of studies indexed in two or more databases) was calculated before the screening. Nine hundred ninety documents were obtained using our search terms and ancestral citation searching.

We retrieved documents published as journal articles, conference proceedings, or book chapters, where most metadiscourse and L2 writing studies are published. One hundred thirty-three records were removed during the screening process because they were unrelated to our focus on metadiscourse. Seven non-empirical studies, which include book reviews and position papers, were excluded. Ten studies were removed because they investigated student writing in other L2s or L1 English rather than L2 English, despite our utilising of contextual search terms. One record was removed for being the wrong document type. One study of L3 English and four that did not specify participants' language backgrounds were excluded. Empirical studies which addressed writing by non-native English researchers and ESL adults and the ones which addressed the writing of textbooks, or students' spoken presentations, were excluded ($N = 32$). Six other empirical studies were excluded as they addressed L2 English student writing. The main focus of these six studies was not voice but interrater reliability, writing quality, writing assessment, Coh-Metrix measurement, or students' use of information. The term voice did appear in these studies, but it was not employed as their main concern.

An important consideration for our exclusion criteria was the theoretical framework. Our focus on students' voice refers to voice defined by metadiscourse models. Therefore, to be cautious, empirical studies that focused on other theoretical frameworks or approaches, whether mixed with the metadiscourse model or not, were removed in the exclusion process ($N = 26$). The aforementioned theoretical frameworks and approaches mainly included Bakhtin's dialogism, Appraisal Theory, Complex Dynamic Systems Theory (CDST), Conversational Analysis, Critical Discourse Analysis, Multifactorial Approach, and Multidimensional Analysis. Another seven empirical studies were excluded because they were conducted without consideration of metadiscoursal properties.

It is important to note that the decision to exclude studies that mixed the metadiscourse model with other theoretical frameworks or approaches was likely based on the review's specific focus. We wanted to isolate and closely examine the impact of the metadiscourse model on L2 English writing without the influence of other theories and frameworks.

While a longitudinal study that uses CDST and the metadiscourse model to examine the development of voice in students' writing could potentially provide valuable insights, it may not have fit the specific inclusion criteria for this review. We believe that integrating these different theoretical frameworks could have complicated the analysis



*Unrelated (133), not an empirical study (7), not L2 student English writing (10), not specify students' language backgrounds or investigated L3 English (5), not student writing (e.g., writing by L2 English researchers, writing by ESL adults, writing by journal editors; textbooks, students' oral practice) (32), wrong document type (e.g., dissertation) (1), employed a different theoretical framework or approach (e.g., Bakhtin's dialogism, Complex Dynamic Systems Theory, Local Grammar, information structure, thematic organization, Appraisal Theory, multifactorial approach, multidimensional analysis, Conversational Analysis, Critical Discourse Analysis, Walková's model of self-mention) (27), not focus on students' use of voice (e.g., information use, interrater reliability, writing quality, writing assessment, Coh-Metrix measurement) (6), not consider metadiscourse properties (7)
 ** not focus on students' use of voice (e.g., writing assessment, human rating, feedback, instruction) (8), employed a different theoretical framework (e.g., Critical Discourse Analysis, Appraisal Theory, Bakhtin's dialogism) (4), not distinguish L2 student writing from writing by L2 researchers in the data (1), not consider metadiscourse properties (1), not an empirical study (1), mixed L2 writing with L1 writing in the data (3)

Fig. 1. PRISMA diagram.

and interpretation of the results. Excluding certain studies does not necessarily mean they are lower quality or less important. It simply means we had specific inclusion and exclusion criteria and research questions or areas of interest.

A further assessment of the screened studies was performed in the PRISMA process to guarantee an accurate and eligible filtering result for this review. As shown in Fig. 1, eight empirical studies were excluded because they used voice as an angle to address their primary concerns, such as writing assessment, human rating, feedback, or teacher's instruction. One more study was removed for adopting Appraisal Theory, one for following an analytical framework of Critical Discourse Analysis, another for adopting Bakhtin's dialogism, and an additional exclusion was made due to their disinterest in metadiscourse properties.

In addition to the above exclusion, another empirical study was excluded because its data included writing by L2 English researchers and L2 English students. However, the researchers did not distinguish between the two in their analysis. Similarly, three studies were excluded for mixing data collected from native and non-native English speakers without explanation. As a result of the screening and filtering, 59 studies, based on our specific selection criteria, were included in the systematic review.

2.4. Data coding and analysis

The present review adopts quantitative data collection and analysis methods in the data coding process. It utilised a deductive analytical scheme to meticulously probe into the conceptual and methodological features of the filtered empirical studies. Following the examples set by the classic and widely used metadiscourse models (Crismore et al., 1993; Hyland, 2005a, 2005d, 2005b; Adel, 2006), other influential empirical studies in metadiscourse (Adel, 2010, 2010del and Mauranen, 2010;

Dahl, 2004; Mauranen, 1993; Hyland and Milton, 1997; Hyland and Tse, 2004), and the examples of systematic reviews (Plonsky and Gass, 2011; Aull and Lancaster, 2014; Hiver et al., 2021) and methodological reviews in metadiscourse and second language acquisition (Crismore and Abdollahzadeh, 2010; Plonsky and Gass, 2011), the deductive analytical scheme used in this study was developed inductively and recursively by identifying values, variables, and categories of research focus in respect of the above examples. The pilot coding scheme was developed based on 30 empirical voice and L2 English writing studies. Revisions were made based on the draft scheme concerning variables not identified in the initial literature review.

We developed the scheme by focusing on six categories in response to our research questions: research designs, data sources, conceptual considerations (RQ1), contexts, participants' language backgrounds, texts/corpora, and reporting practices (RQ2).

Each category is analysed through a set of variables. Variables such as the corpus size are open-ended. In contrast, other variables contain categorical values, such as metadiscourse framework and analytical software. Most values can be identified in the Method, Results and Discussion sections. However, variables need to be interpreted when the target study involves analytical software, more than one corpus was used to make a comparison, or the data were collected from multiple groups of participants. Therefore, the coding was cross-checked in the recoding procedure to verify the accuracy of all the variables included in this review.

The two authors coded all variables in this paper to ensure inter-coder reliability. Data were coded independently by the two coders. Then the coding results were compared and discussed. The differences in variables were resolved, and the refined variables were recorded. Using Cohen's Kappa coefficient to determine the consistency between coders, the result was 0.885. It indicates a high level of agreement between the

coders, according to Landis and Koch (1977) and suggests that the coding is reliable. The findings will be presented in the form of frequency counts and proportions.

3. Results and Discussion

3.1. Theoretical frameworks

To continue discussing how the forms and frequencies of metadiscourse markers were identified, we analysed the theoretical frameworks adopted by the reviewed studies regarding metadiscourse markers and their taxonomies. Our review focused on voice in metadiscourse; hence, only markers connected to voice are listed. We followed the terms used by the researchers since many of the reviewed studies adopted a modified framework or taxonomy, and the terminology might be inconsistent (See Table 1 below).

As can be seen from Table 1, many of the reviewed studies investigated one or more sub-types of interactional markers. 45.8% of them (n = 27) investigated self-mention markers, 38.9% (n = 23) investigated hedges, 35.6% (n = 21) investigated boosters, 37.3% (n = 22) investigated attitude markers, 16.9% (n = 10) investigated engagement markers. Only a few studies reported more detailed sub-classification of the targeted markers. For example, Lee and Deakin (2016) analysed each sub-category of hedges, boosters, attitude markers, self-mention and engagement markers used in argumentative writing. Besides this, our review found that many researchers would use more personalised terminology rather than a ready-made taxonomy list when investigating voice markers.

Zhang and Zhang (2021b) adopted Hyland's (2005b) interactional model in writing instructions, not using an existing model to prepare a taxonomy in their research. Among other reviewed studies, Hyland's metadiscourse model or taxonomy (Hyland, 1996, 1998a, 1998b, 2002, 2004, 2005a, 2005b, 2008, 2010, 2012; Hyland and Tse, 2004) appeared to be the prominent adoption, used by 40 studies. Other less commonly used models or taxonomies were Lehman (2018), vaniĉ & Camps (2001), Ivaniĉ (1998), and others. Unfortunately, with a close examination, the three reviewed studies did not state which previous model or taxonomy they used as a reference in their analysis.

3.1.1. Text, corpora and context

We analysed the adopted corpora and sub-corpora, collected texts, corpus size, language represented in corpora, and the study context in these corpus-based or corpus-assisted studies, as shown in Table 2.

As shown in Table 2, 44% of the reviewed studies (n = 26) investigated undergraduate students' writing or BA thesis. In comparison, 11.9% (n = 7) investigated master's thesis or writing at the postgraduate level, and 5.0% (n = 3) investigated doctoral dissertations. Only two of the reviewed studies recruited participants from high school. Undergraduate students seemed to be the largest source of participants in L2 English writing studies due to ease of access.

More than half of the reviewed studies (52.5%) collected writing from participants with the same L1 background. The remaining studies, 47.4% (n = 28), were conducted based on writing by L2 English users with diversified L1 backgrounds. Moreover, argumentative writing and academic writing were the two prominent choices of genres in our pool of studies, 45.8% and 40.7%, respectively.

3.1.2. Reporting practices

Based on our research design and study types analysis, Table 3 represents the reviewed studies' reporting practices and statistical analyses.

Our review of reporting practices revealed that most studies (84.7%) reported descriptive statistics; the others were either case studies or qualitative studies. The inferential statistics were employed in the reviewed studies: 8.0% (n = 5) used Chi-square, and another 8.0% used the Mann-Whitney U test, the percentage of which was higher than other statistical tests.

Table 1
Metadiscourse Models and Taxonomies of voice.

Metadiscourse Taxonomies of Voice	n	Percentage
Stance		
stance markers (as a whole)	1	1.6%
self-mentioning pronouns/ self-mention markers/ self-mentions/ self-reference/person markers	27	45.8%
pronoun mentions	1	1.6%
exclusive pronouns	1	1.6%
authorial self-mention	3	5.0%
Hedges		
hedges/hedging (as a whole)	23	38.9%
content-oriented	1	1.6%
accuracy-oriented	1	1.6%
writer-oriented	1	1.6%
reader-oriented	1	1.6%
modal auxiliaries, epistemic lexical verbs, epistemic adjectives and adverbs	1	1.6%
Boosters		
boosters/boosting (as a whole)	21	35.6%
Emphatics	2	3.0%
amplifying adverbs	1	1.6%
modal auxiliaries, epistemic lexical verbs, epistemic adjectives and adverbs	1	1.6%
Attitude Markers		
attitude markers (as a whole)	22	37.3%
attitude verbs	1	1.6%
sentence adverbs	1	1.6%
Adjectives	1	1.6%
central point articulation	1	1.6%
Engagement		
engagement markers (as a whole)	10	16.9%
reader pronouns/reader reference	5	8.0%
Directives	5	8.0%
Questions	1	1.6%
Imperatives	2	3.0%
obligation modals	1	1.6%
personal aside	1	1.6%
shared knowledge	1	1.6%
rhetorical questions	1	1.6%
interactional voice	1	1.6%
Other		
generalisation markers	1	1.6%
personal positioning	1	1.6%
personal pronouns	1	1.6%
connectors, adverbs, prepositions	1	1.6%
explicit and implicit authorial references	1	1.6%
quotation: quotation fragment brief quotation, extended quotation	1	1.6%
citation: attribution, links between sources, evaluation	1	1.6%
citation: autobiographical self, discorsal self, authorial self	1	1.6%
autobiographical self, discorsal self, authorial self	2	3.0%
individual self, collective self, depersonalised self	2	3.0%
personal pronouns, generic pronouns, impersonal structures	1	1.6%
writer voice: nominalisation	1	1.6%
writer voice: personal voice	1	1.6%
author's voice: superlatives, adjectives, adverbs, Semi-modal verbs, Pronouns	1	1.6%
author's voice: assertiveness, reiteration of a central point, self- identification, authorial presence and autonomy of thought	1	1.6%
self-reflection	1	1.6%
self-representation	1	1.6%
stance construct	1	1.6%
generalisation markers	1	1.6%
metadiscourse: marked as belonging to another text, paraphrase	1	1.6%
invitational and directive	1	1.6%
Metadiscourse Models Adopted		
Hyland (1996)	1	1.6%
Hyland (1998a)	1	1.6%
Hyland (1998b)	1	1.6%
Hyland (2002)	3	5.0%
Hyland (2004)	2	3.0%
Hyland (2005a)	10	16.9%
Hyland (2005b)	8	13.6%
Hyland (2008)	7	11.9%

(continued on next page)

Table 1 (continued)

Metadiscourse Taxonomies of Voice	n	Percentage
Hyland (2010)	2	3.0%
Hyland (2012)	2	3.0%
Hyland and Tse (2004)	3	5.0%
Aull et al. (2017)	1	1.6%
Borg (2000)	1	1.6%
Burgess and Ivanić (2010)	1	1.6%
Chang and Tsai (2014)	1	1.6%
Ivanić (1998)	2	3.0%
Ivanić and Camps (2001)	3	5.0%
Crompton (1997)	1	1.6%
Fairclough (1992)	1	1.6%
Lee and Deakin (2016)	1	1.6%
Lehman (2018)	3	5.0%
Maley and Kiss (2018)	1	1.6%
Maton (2007)	1	1.6%
Matsuda (2001)	1	1.6%
Reilly et al. (2005)	1	1.6%
Swales (1990)	1	1.6%
Swales (1990), Petrić (2007)	1	1.6%
Hyland (2002), Clark and Ivanić (1997a), Tang and John (1999) Karahan (2013)	1	1.6%
Hyland (2005a), Kondowe (2014)	1	1.6%
Hyland and Milton (1997)	1	1.6%
Aull & Lancaster. (2014), Hyland (2005b), Quirk et al. (1985)	1	1.6%
Unclear	3	5.0%

We were concerned that 22% of the reviewed studies (n = 13) did not report the tokens of corpora or sub-corpora, and 31% (n = 18) were unclear whether the metadiscourse function of voice markers was checked or confirmed in the co-text. More importantly, 49% (n = 29) of the studies did not report the reliability of coding or annotation in their analysis. These findings suggest that the reporting practices in this field need to be more thorough to improve research quality.

3.2. Methodological characteristics

3.2.1. Methodological concern

Methodology in theoretical and applied linguistics is “concerned with the relationship between theory and data” (Ender et al., 2012, p.2); metadiscourse studies are no exceptions. The notion and research practices of metadiscourse are deeply rooted in pragmatics (Vande Kopple, 1985; Hyland, 2005a), which studies language in use, language and its context, the relation between context and language meaning, and communication among certain language speakers. Metadiscourse developed based upon the written discourse; it concerns itself mainly with how readers can understand the organisation and connection of a text, how they can correctly understand the meaning of the writer, how writers can express themselves clearly or what writers think of certain propositions (Vande Kopple, 1985:83–85). It is typically suitable for adopting a qualitative method. In addition, the ability to use lexical items of metadiscourse was considered an indicator of L2 learners’ pragmatic competence in writing (Hyland and Milton, 1997, p. 184).

Per the pragmatic and discursal tradition of metadiscourse studies and the technology used in modern linguistics, analysing metadiscourse markers or patterns by means of concordance analysis is an important qualitative method in these studies (Hyland, 2005a: 198). In addition, a mixed-method study is equally important since qualitative analysis is usually conducted after data collection and annotation. Moreover, as Hyland (2005a) pointed out, interviews are necessary in certain cases to provide explanations from the perspective of participants other than researchers. Thus, a triangulation of corpus data with interviews or feedback is often adopted in metadiscourse studies.

Turning to research design characteristics (Fig. 2 & Table 4), we looked at the general research design methods and the approaches adopted in the reviewed studies. Our analysis found that among the examined empirical studies of voice and L2 English writing, more than half of the studies adopted a mixed method. Fig. 2 further shows that the

Table 2

Corpora characteristics and study contexts.

Corpora Characteristics	Study Context	n	Percentage
Corpora Languages	• L2 English	59	100%
	• Native English	6	10%
	• Turkish	4	7.0%
	• Spanish	1	1.6%
Native Language Backgrounds	Single L1 background in one study	31	52.5%
	Diversified L1s in one study	28	47.4%
Corpus Size	Quantity of texts		
	1–10	4	7.0%
	11–30	2	3.0%
	30–50	14	23.7%
	51–100	20	33.9%
	101–200	6	10.2%
	201–500	10	16.9%
	Unreported	3	5.0%
	Average text length (words)		
	200<	5	8.0%
	201–300	15	25.4%
	310–500	29	49.2%
	500–1000	4	7.0%
1000–2000	1	1.6%	
Unreported	10	16.9%	
Corpus Time Frame	1<	35	59.3%
	2–4	9	15.3%
	4>	1	1.6%
	5>	1	1.6%
	Unreported	13	22%
Writing Genres Represented	Argumentative writing	27	45.8%
	Students’ academic writing	20	33.9%
	Expert academic writing	4	7.0%
	Sections of academic writing selected for study aim		
	Full text	1	1.6%
	Abstract	2	3.0%
	Introduction	1	1.6%
	Result	4	7.0%
	Discussion	6	10.2%
Academic Writing Disciplines	• English	1	1.6%
	• Applied linguistics	2	3.0%
	• Psychology	2	3.0%
	• Education	1	1.6%
	• Business	2	3.0%
	• Expository writing	2	3.0%
	Other		
	• Autobiography journal	1	1.6%
	• Fiction	1	1.6%
	• Application Essay	1	1.6%
	• Letter	1	1.6%
	• Poetry	1	1.6%
	• Colloquial Communication	1	1.6%
Study Contexts	Institutionally-based participants	58	98.3%
	Institutionally-based corpora	52	88%
	Texts obtained from one or more institutions	18	30.5%
	Texts obtained from a literature data repository	20	33.9%
	Texts obtained from an English language test repository	4	7.0%
	Unreported	1	1.6%
	Institutional Settings	University	56
High School		2	3.0%
Unreported		1	1.6%
Participants’ Academic Degrees/Courses	Undergraduate	26	44%
	M.A.	7	11.9%
	Ph.D.	3	5.0%
	EFL or ESL courses	16	27.1%
	English language exams	5	8.0%

choice of method was split across quantitative (15.2%), qualitative (32.2%), and mixed methods studies (54.2%). Our result showed that this prominent adoption of mixed methods might mirror growing trends in metadiscourse and second language acquisition that value multiple methods as productive and integrated.

Over one-third of the studies (38.9%) prominently triangulated text

Table 3
Reporting practices and statistical analyses.

Analysis and Reporting of the Reviewed Studies	Statistical Analyses	n	Percentage
Analytical Tests	Descriptive statistics	50	84.7%
	Inferential statistics		
	Chi-square	5	8.0%
	Mann-Whitney <i>U</i> test	5	8.0%
	Kruskal-Wallis (one-way) test	4	7.0%
	Wilcoxon signed-rank test	2	3.0%
	ANOVA	2	3.0%
	t-test (paired-samples/dependent)	4	7.0%
	Post-hoc analysis	3	5.0%
	Log-likelihood	1	1.6%
	Effect size	1	1.6%
	Person correlation coefficient	1	1.6%
	Correlation coefficient	2	3.0%
	Alpha coefficient	1	1.6%
	Correlation	2	3.0%
	Multiple regression analysis	1	1.6%
	Multiple correspondence analysis	1	1.6%
	Differential item functioning	1	1.6%
	Maximum likelihood estimation	1	1.6%
	Confirmatory factor analysis	2	3.0%
	Exploratory factor analysis	1	1.6%
	Mixed-effects analysis	1	1.6%
	Spearman's rho correlation coefficient	1	1.6%
	Principal component analysis	2	3.0%
	Correlation matrix	1	1.6%
	Regression analysis	2	3.0%
	Bonferroni adjustment	1	1.6%
Reporting total tokens of the corpora/sub-corpora	Yes	46	78%
	No	13	22%
Reporting metadiscourse marker types	Full quantitative results	54	92%
	Most frequent markers	5	8.0%
Checking the metadiscourse function in the co-text	Yes	37	62.7%
	No	4	7.0%
	Unclear	18	31%
Reporting reliability of coding/annotation	Yes		
	Inter-rater reliability	12	20.3%
	Double coding/annotation of part of the markers	10	16.9%
	Double coding/annotation of all markers	3	5.0%
	External expert	5	8.0%
	No	29	49%

analysis with analytical approaches such as interviews, feedback, peer comments, questionnaires, instructions, writing interventions, written narratives, and quizzes. 10% of the studies (six studies) compared the

use of voice at different proficiency levels, and another 10% compared L2 English and native English. 6.8% of the studies (four studies) compared L2 students writing with expert writing. 5% of the studies (three studies) were cross-disciplinary studies of students' academic writing. Two reviewed studies compared markers in different sections of academic writing; one was cross-register analysis, and one was cross-gender analysis. One of our reviewed studies analysed the language backgrounds of the participants, and one study analysed the writers' backgrounds, which were scarce in this field. In addition, two reviewed studies were longitudinal, and one adopted a quasi-experimental design.

Some studies in the pool ($N = 59$) adopted more than one approach to investigate students' use of voice markers. For instance, Wu and Buripakdi (2021) approached writers' identity and its construction in EFL doctoral dissertations by triangulating corpus data with three types of analyses, i.e., analyses of a questionnaire, written narratives and interviews. Another example is Yoon (2021), who investigated authorial voice in L2 English learner writing. The researcher conducted two comparisons: across language proficiency levels; and three groups of L1 backgrounds. These studies provided valuable insights into the development of voice, metadiscourse and L2 English writing by relating corpus data and textual analysis with more comprehensive variables.

3.2.2. Study types

The reviewed studies were further classified into ten discrete

Table 4
Research design.

Research Design	n	Percentage
<i>Method</i>		
Quantitative	9	15.2%
Qualitative	19	32.2%
Mixed	32	54.2%
<i>Approach</i>		
Cross-sectional	2	3.0%
Cross-disciplinary	3	5.0%
Cross-linguistic	5	8.0%
Cross-language-background	1	1.6%
Cross-proficiency-level	6	10%
Cross-register	1	1.6%
Cross-gender	1	1.6%
Cross-writer-background	1	1.6%
L2 English vs. L1 English	6	10%
L2 students vs. expert writers	4	6.8%
Triangulated with interview/feedback/peer comments/questionnaire/instruction/writing intervention/written narratives/quiz/narrative inquiry	23	38.9%
Longitudinal	2	3.0%
Quasi-experimental	1	1.6%

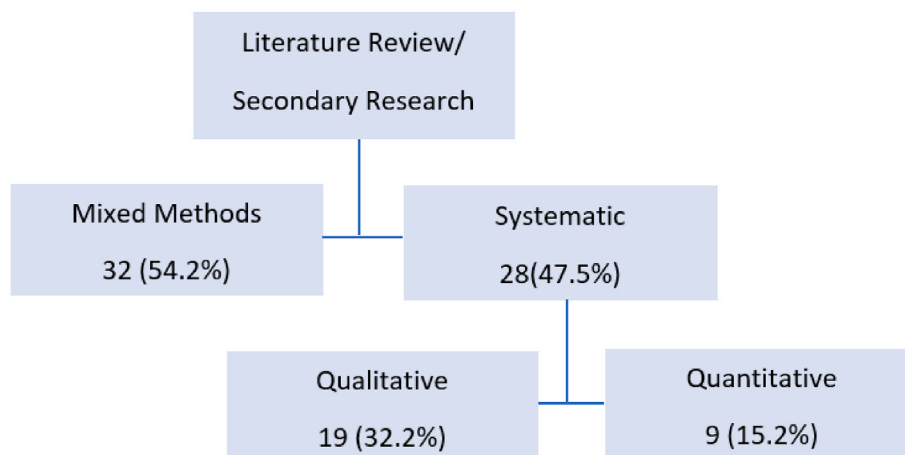


Fig. 2. Major types of secondary research.

categories based on the results presented in Table 4 above. Classification of the reviewed study types is shown in Table 5. More than one-third of the studies (38.9%) were triangulating in nature. 22% of the reviewed studies ($n = 13$) triangulated statistical data with qualitative analytical strategies, and 16.9% ($n = 10$) triangulated different qualitative analysis approaches. The main reason might be that most reviewed studies used corpus as a data source, making triangulation more natural and easy.

Although contrastive studies are commonly seen in metadiscourse and learner corpus studies, less than half were conducted from a contrastive perspective among our reviewed studies. Among studies in the pool ($N = 59$), 20.3% compared the use of voice either in L2 English writing and L1 English writing ($n = 6$) or compared that in three groups, i.e., participants' native language group, L2 English group and native English group ($n = 5$). Only one stood out because participants' native languages were used as variables in their analysis (Yoon, 2021).

Other reviewed studies that take a contrastive perspective include eight studies. Five of them were L2 English academic writing studies, two cross-sectional studies and three cross-disciplinary studies. The other three contrastive studies were one cross-register study comparing L2 English learner's writing and colloquial communication, one cross-gender study comparing voice markers in writing by male and female L2 English learners, and one study using writer background information as variables to investigate voice construction in L2 English writing as one of its analytical procedures (Zhao, 2019).

Among the 16.9% of reviewed studies ($n = 10$), two studies compared proficiency levels of learner English and used CEFR as referencing criteria to divide their collected texts into proficiency groups (Bax et al., 2019; Yoon, 2021). Two studies used criteria set by the participants' universities when assessing proficiency levels of L2 English learner writing (Petrić, 2012; Abbuhl, 2012). One study's referencing criteria included TOFEL's criteria and criteria set by two learner corpora (Lee and Deakin, 2016). Only one study used grading criteria set by the researchers. Furthermore, two studies which compared L2 English student writing with expert writing were included in this category. They both provided reasonable explanations for considering L2 English student writing as lower proficient and expert writing as higher proficient (Chen and Zhang, 2017; Aull et al., 2017).

Taking L2 English writing as the single focus remained prominent in the reviewed pool of studies, except for 20.3% of the aforementioned studies. This finding may indicate that exploring L2 English writing is adequate for its own characteristics and for spotlighting L2 writing development, using voice and other metadiscourse markers.

3.2.3. Data sources and retrieval

As for the data sources used in the study pool ($N = 59$), the tools used in data retrieval and how they were used, Table 6 shows the details. The classification of data sources overlapped because more than half of the study pool adopted mixed methods (see Table 4). Our review found a prominent adoption (71.2%) of the studies ($n = 42$) created a specialised corpus for their study aim. Adopting a specialised corpus was apparent and reflected specific design requirements, notably in corpus-based metadiscourse studies. 11.9% of the reviewed studies ($n = 7$) collected texts for case studies or qualitative analysis but did not build the collected texts into a small-scaled corpus. 10.2% of the studies ($n = 6$) used existing large-size corpora and had better representativeness to draw on research questions related to groups of language users in a broader sense. For example, Aull et al. (2017) used The Michigan Corpus of Upper-Level Student Papers (MICUSP), and Yoon (2021) used the International Corpus Network of Asian Learners of English (ICNALE); both were deemed suitable corpora for their study aim. Another example that should be mentioned in this review is Çandarlı et al. (2015), a cross-linguistic study which utilised a sub-corpus of the Louvain Corpus of Native English Essays (LOCNESS) to investigate native students' English as well as a self-built corpus of L2 English and a self-built corpus of native Turkish.

Besides using existing corpora and creating a specialised corpus that

Table 5
Categories of study types.

Category	Sub-category	Example Studies	n	Percentage
Triangulation approaches in the studies of English learner writing and voice	1) Triangulating quantitative data with qualitative analytical strategies	Petrić (2012)	23	38.9%
	2) Triangulating different approaches in qualitative analysis	Lehman and Anderson (2021), Juliaty (2019)	13	22%
Adopting quasi-experimental design	Adopting both a quasi-experimental design and a triangulation approach	Zhang and Zhang (2021a)	1	1.6%
Cross-sectional studies in students' academic writing	Comparing the use of voice in different sections of dissertations and research articles	Ngampradit (2020)	2	3.0%
Cross-disciplinary studies in students' academic writing	Comparing the use of voice in L2 English academic writing of different disciplines	Yasmin and Mahmood (2017)	3	5.0%
Comparison of the use of voice in different languages or language variations			12	20.3%
	1) Cross-linguistic studies among participants' native language, L2 English and native English	Lorés-Sanz (2011), Çandarlı et al. (2015)	5	8.0%
	2) Comparison between L2 English and L1 English	Can and Cangır (2019), Chen and Nassaji (2015)	6	10%
	3) Using participants' native language as a variable	Yoon (2021)	1	1.6%
Comparison of the use of voice in writings by participants of different proficiency levels			10	16.9%
	1-1) Comparison among L2 student groups by using the Common European Framework of Reference (CEFR)	Bax et al. (2019), Yoon (2021)	2	3.0%
	1-2) Comparison among L2 student groups by using criteria set by English language test or learner corpora	Lee and Deakin (2016)	1	1.6%
	1-3) Comparison among L2 student groups by using criteria set by the participants' university	Petrić (2012), Abbuhl (2012)	2	3.0%
	1-4) Comparison among L2 student groups by using other criteria	Wu and Paltridge (2021)	1	1.6%
	2) Comparison between the writing by L2 students and expert writers	Chen and Zhang (2017), Aull et al. (2017)	4	6.8%

(continued on next page)

Table 5 (continued)

Category	Sub-category	Example Studies	n	Percentage
Cross-register	Using registers as the within-subject covariates to investigate the use of voice by L2 English learners	Qin and Uccelli (2019)	1	1.6%
Cross-gender	Comparing voicing markers in writing by male and female L2 English learners	Abdul Aziz et al. (2016)	1	1.6%
Cross-writer-background	Using writer background information as variables to investigate voice construction in L2 English writing	Zhao (2019)	1	1.6%
Adopting a longitudinal perspective in the study of students' use of voice	1) Using an existing diachronic learner corpus	Liardét (2018)	1	1.6%
	2) Compiling a longitudinal learner corpus by the researchers	Crosthwaite and Jiang (2017)	1	1.6%

is only for one-time use, we found that 5.0% of the studies (n = 3) used corpora that the researchers themselves built. These corpora could be repeatedly utilised in empirical studies. For example, [Liardét \(2018\)](#) used the Chinese Longitudinal Learner Corpus (CLLC) created by the researcher in 2013. In another example, [Lee and Deakin \(2016\)](#) utilised a sub-corpus of the Corpus of Ohio Learner and Teacher English (COLTE), a 5-year corpus project of ESL created by the researchers' unit at their university. These studies showed the value and advantage of building corpora for long-term interests, which are much more beneficial and practical for the researchers and their peers than building a "disposable" corpus only for one study.

Most reviewed studies used one of the corpus software tools, among which the most popular tool was AntConc ([Anthony, 2022](#)), utilised by 23.7% of the studies (n = 14). 11.9% of the studies (n = 7) used WordSmith Tools. [Çandarlı et al. \(2015\)](#) was the only study utilising two software, UAM CorpusTool and AntConc. Other software tools used by the studies include UAM CorpusTool (n = 4), Authorial Voice Analyzer (n = 2), Lancsbox (n = 1), RegExr (n = 1), SiNLP, Text Inspector ([Bax, 2012](#)) (n = 1). Lancsbox ([Brezina et al., 2015, 2020](#)) is a relatively new corpus linguistics software that shares similarities with AntConc and WordSmith Tools and contains a function called *GraphColl* which allows for building tailored collocation networks. Authorial Voice Analyzer ([Yoon, 2017a, 2017b, 2021](#)) stood out among the above tools for representing the new trend in combining the metadiscourse framework with corpus linguistics and NLP.

Additionally, as seen in [Table 6](#), NVivo and ATLAS.ti were used by a few studies for the convenience of coding, 5.0% (n = 3) and 1.6% (n = 1), respectively.

Unsurprisingly, 22% of the reviewed studies (n = 13) did not use software tools because they investigated a small size of texts and manual annotation or coding was acceptable in terms of efficiency. However, another 22% of the reviewed studies (n = 13) did not report using software tools in the annotating procedure.

One problem with how the metadiscourse markers were annotated or coded, either by software tools or manually, was found in our review. Most studies in the pool adopted a predetermined taxonomy or a modified taxonomy based on previous literature to annotate metadiscourse markers, except that only 3.0% of the studies (n = 2) identified metadiscourse markers intelligently. However, only 22% of the studies (n = 13) reported the procedure and result of manual checking after the

Table 6

Data sources and corpus tools.

Data Sources	Corpora/Non-Corpora Data	n	Percentage	
Corpus Data	- Using existing corpora	6	10.2%	
	- Creating a specialised corpus for the study aim	42	71.2%	
	- Using an existing corpus created by the researcher(s)	3	5.0%	
	- Texts collected in case studies or qualitative analysis	7	11.9%	
Non-Corpus Data	• interview	13	22%	
	• feedback	3	5.0%	
	• peer comments	2	3.0%	
	• questionnaire	4	7.0%	
	• instruction	1	1.6%	
	• writing intervention	2	3.0%	
	• written narratives	2	3.0%	
	• quiz	2	3.0%	
	• narrative inquiry	1	1.6%	
Software Tools	Corpus/Other Tools	n	Percentage	
Corpus Software Tools	AntConc	14	23.7%	
	WordSmith Tools	7	11.9%	
	UAM CorpusTool	4	7.0%	
	Authorial Voice Analyzer	2	3.0%	
	Lancsbox	1	1.6%	
	RegExr	1	1.6%	
	SiNLP	1	1.6%	
	Text Inspector	1	1.6%	
	No corpus software tools used	13	22%	
	Unreported	13	22%	
	Other Software Tools	NVivo	3	5.0%
		ATLAS.ti	1	1.6%
	Purposes of Corpus Software Tools	Identify forms and frequencies of metadiscourse markers from a predetermined taxonomy or with modification	25	42.4%
Identify metadiscourse markers using a tag query		2	3.0%	
Intelligently identify metadiscourse markers		2	3.0%	
Generate concordance lines for qualitative analysis		12	20.3%	
Manually identify/code metadiscourse markers by following a predetermined taxonomy or with modification		11	18.6%	
Manually and intelligently identify metadiscourse markers		4	7.0%	
Identify forms and/or frequencies of metadiscourse markers from a predetermined taxonomy or with modification but not report by using tools or manually		10	16.9%	
Manually identify/code metadiscourse markers during the pilot study		1	1.6%	
Manually check the results after annotation/coding by tools		13	22%	
Unreported		4	7.0%	

annotation. Furthermore, only one study (1.6%) reported that the researchers conducted a pilot study for intercoder analysis and evaluation in their analytic framework before analysing the whole corpus ([Akbas and Hardman, 2017](#)). This problem raised our concern because a pilot study is necessary for ensuring the feasibility and validity of research, especially when large-scale data is about to be utilised or in comparative studies of language variations, genres or registers, proficiency levels, language backgrounds, and genders.

4. Conclusion

This concluding section highlights the conclusion and suggestions for further research. This systematic review responded to how L2 English students' use of voice could be investigated by analysing 59 empirical studies in this field. The unique intellectual contribution of the current

study lies in its answers to the two specific research questions that guided our study. The first question reviewed the metadiscourse theoretical frameworks adopted in the reviewed empirical studies of voice in L2 English writing. The answer to this question showed the commonly-used metadiscourse theoretical frameworks used in L2 writing research with a special emphasis on voice. Consequently, future researchers can build on this result and develop a more refined and nuanced understanding of metadiscourse and voice in L2 English writing. Our results could inform other L2 contexts other than English.

On the other hand, the second research question focused on reviewing the methodological characteristics of the reviewed empirical studies of voice in L2 English writing. Our answer to the second research question provided a more comprehensive understanding of the current research on voice in L2 writing. It contributed to the rigour and quality of research on voice in L2 writing. Also, our answer to this research question enabled us to identify common approaches, best practices, and areas where more research is needed. Moreover, we identified potential limitations and sources of bias in the existing literature. This can help guide future research and improve the quality of studies on this topic, inform the future research design and improve the validity and reliability of findings.

Two main strengths and three weaknesses were identified in this review. Methodological triangulation is the first strength shown by our review. Studies of voice and Second Language Acquisition (SLA) proved that triangulating interviews, questionnaires, or feedback can provide supportive evidence to corpus data. The other strength is the utilisation and development of software and NLP tools. 'Authorial Voice Analyzer' as an NLP tool combines a metadiscourse framework and large-scale corpus to provide quantitative results in writing studies and benefit all researchers interested in voice studies. However, three weaknesses were found in our review. Some studies were found to be weak in their methodological practices, report of corpus data sources, and research of L2 English learners' backgrounds in writing.

Appendix. (1)

List of Search Terms

meta discourse	AND	Writing
Voice		L2 writing
Meta discourse marker		non-native writing
Metatext		student writing
textual voice		student essay
voice structure		student text
voice strength		undergraduate writing
voice construction		graduate writing
writ* identity		academic writing
writer's voice		argumentative writing
writers' authorial positioning		persuasive writing
author's voice		narrative writing
individual voice		descriptive writing
Stance		expository writing
stance marker*		EFL writing
writer* stance		EFL context
authorial stance		ESL writing
epistemic stance		EAP writing
interaction*		English
hedg*		English writing
booster*		English essay
Attitude		Learner
attitude markers		Dissertation*
Engagement		Thesis*
self-mention		Master
self-reference		Doctoral
self-referential pronouns		composition*
Quotation		
Citation		

(continued on next page)

4.1. Suggestions for future research

Based on the above findings, we provide suggestions for future research on L2 English students' use of voice. Our findings suggest that research will be better served by improving certain methodological practices, such as reporting tools' reliability, annotation and coding reliability. Randomly assigning participants will provide more reliable results when the research is conducted by adopting an experimental design. Secondly, reporting corpus design and corpus data also need to be improved to provide the full context of the research and persuade readers with interpretability and reliability. Furthermore, based on students' prior education, and cultural or social experiences, L2 English students from different backgrounds will better understand the appropriate uses of voice in L2 writing. Thus, more rigorous studies about L2 English learners' background characteristics will provide a better view of voice and SLA. Finally, a lack of longitudinal studies found in our review suggests the need for more diachronic research design in this field. Looking forward, the findings of this review should prompt the field of voice and SLA to reflect on and further investigate how voice studies are carried out and reported. Ideally, this study provided insights for voice studies in a broader scope, covering more types of language variations and registers.

Declaration of competing interest

There are no conflicts of interest with any other parties.

Acknowledgements

The authors would like to thank the editor of *Ampersand: An Interdisciplinary Journal of Language Sciences and Bilingualism* and the anonymous reviewers for their constructive feedback during the review process. Open Access funding of this article is provided by Qatar National Library (QNL), Doha, Qatar.

(continued)

 Adverbs
 linking adverbial
 Adjectives
 averral*
 attribution*
 phrase*
 clause*

References

- Abuhl, R., 2012. Using self-referential pronouns in writing: the effect of explicit instruction on L2 writers at two levels of proficiency. *Lang. Teach. Res.* 16 (4), 501–518.
- Abdul Aziz, R., Chin, C.J., Nordin, N.M., 2016. The use of interactional metadiscourse in the construction of gender identities among Malaysian ESL learners. *3L: Southeast Asian J. Engl. Lang. Stud.* 22 (1), 207–220.
- Ädel, A., 2006. *Metadiscourse in L1 and L2 English*. John Benjamins Publishing Company.
- Ädel, A., 2010. Just to give you kind of a map of where we are going: a taxonomy of metadiscourse in spoken and written academic English. *Nordic J. Engl. Stud.* 9 (2), 69–97.
- Ädel, A., Mauranen, A., 2010. Metadiscourse: diverse and divided perspectives. *Nordic J. Engl. Stud.* 9 (2), 1–11.
- Ahmed, A.M., Zhang, X., Rezk, L.M., Pearson, W.S., 2023. Transition Markers in Qatari University Students' Argumentative Writing: A Cross-Linguistic Analysis of L1 Arabic and L2 English, 100110. *Ampersand*.
- Akbas, E., Hardman, J., 2017. An exploratory study on authorial (In)visibility across postgraduate academic writing: dilemma of developing a personal and/or impersonal authorial self. In: Hatipoglu, C., Akbas, E., Bayyurt, Y. (Eds.), *Metadiscourse in Written Genres: Uncovering Textual and Interactional Aspects of Texts*. Peter Lang, pp. 139–174.
- Anthony, L., 2022. *AntConc V. 4.1.4*. [software]. Waseda University, Tokyo, Japan. Available at: <https://www.laurenceanthony.net/software/antconc/>.
- Appel, R., Szeib, A., 2018. Linking adverbials in L2 English academic writing: L1-related differences. *System* 78, 115–129.
- Aull, L.L., Lancaster, Z., 2014. Linguistic markers of stance in early and advanced academic writing: a corpus-based comparison. *Writ. Commun.* 31 (2), 151–183.
- Aull, L.L., Bandarage, D., Miller, M.R., 2017. Generality in student and expert epistemic stance: a corpus analysis of first-year, upper-level, and published academic writing. *J. Engl. Acad. Purp.* 26, 29–41.
- Bax, S., 2012. *Text Inspector*. [software]. Online text analysis tool. Available at: <https://textinspector.com/>.
- Bax, S., Nakatsuhara, F., Waller, D., 2019. Researching L2 writers' use of metadiscourse markers at intermediate and advanced levels. *System* 83, 79–95.
- Borg, E., 2000. Citation practices in academic writing. In: Thompson, P. (Ed.), *Patterns and Perspectives: Insights into EAP Writing Practices*. Reading: University of Reading, pp. 27–45.
- Brezina, V., McEnery, T., Wattam, S., 2015. Collocations in context: a new perspective on collocation networks. *Int. J. Corpus Linguist.* 20 (2), 139–173.
- Brezina, V., Weill-Tessier, P., McEnery, A., 2020. *#LancsBox v. 5.x*. [software]. Available at: <http://corpora.lancs.ac.uk/lancsbox>.
- Burgess, A., Ivanić, R., 2010. Writing and being written: issues of identity across timescales. *Writ. Commun.* 27 (2), 228–255.
- Bychkovska, T., Lee, J.J., 2017. At the same time: lexical bundles in L1 and L2 university student argumentative writing. *J. Engl. Acad. Purp.* 30, 38–52.
- Can, T., Cangir, H., 2019. A corpus-assisted comparative analysis of self-mention markers in doctoral dissertations of literary studies written in Turkey and the UK. *J. Engl. Acad. Purp.* 42, 100796.
- Çandarlı, D., Bayyurt, Y., Martı, L., 2015. Authorial presence in L1 and L2 novice academic writing: cross-linguistic and cross-cultural perspectives. *J. Engl. Acad. Purp.* 20, 192–202.
- Chang, P., Tsai, Chin-Chung, 2014. EFL doctoral students' conceptions of authorial stance in academic knowledge claims and the tie to epistemic beliefs. *Teach. High. Educ.* 19 (5), 525–542.
- Chen, C., Zhang, L.J., 2017. An intercultural analysis of the use of hedging by Chinese and Anglophone academic English writers. *Appl. Ling. Rev.* 8 (1), 1–34.
- Chen, S., Nassaji, H., 2015. Self-mention and identity construction in statement of purpose [special issue for INTE]. *TOJET - Turkish Online J. Educ. Technol.* 476–480.
- Clark, R., Ivanić, R., 1997a. *The Politics of Writing*. Routledge, London.
- Clark, R., Ivanić, R., 1997b. Critical discourse analysis and educational change. In: *Encyclopedia of Language and Education*. Springer, Dordrecht, pp. 217–227.
- Crismore, A., Abdollahzadeh, E., 2010. A review of recent metadiscourse studies: the Iranian context. *Nordic J. Engl. Stud.* 9 (2), 195–219.
- Crismore, A., Markannen, R., Steffensen, M.S., 1993. Metadiscourse in persuasive writing: a study of texts written by American and Finnish university students. *Writ. Commun.* 10 (1), 39–71.
- Crompton, P., 1997. Hedging in academic writing: some theoretical problems. *Engl. Specif. Purp.* 16, 271–287.
- Crosthwaite, P., Jiang, K., 2017. Does EAP affect written L2 academic stance? A longitudinal learner corpus study. *System* 69, 92–107.
- Dahl, T., 2004. Textual metadiscourse in research articles: a marker of national culture or of academic discipline? *J. Pragmat.* 36 (10), 1807–1825.
- Ender, A., Leemann, A., Wälchli, B. (Eds.), 2012. *Methods in Contemporary Linguistics*. Walter de Gruyter.
- Fairclough, N., 1992. *Discourse and Social Change*. Polity Press, Cambridge.
- Flowerdew, J., 2018. *Reading in a Second Language: Moving from Theory to Practice*. Routledge.
- Gong, H., Barlow, M., 2022. A corpus-based analysis of research article macrostructure patterns. *J. Engl. Acad. Purp.* 58, 101138.
- Granger, S., 2015. Contrastive interlanguage analysis: a reappraisal. *Int. J. Learn. Corpus Res.* 1 (1), 7–24.
- Hinkel, E., 2002. *Second Language Writers Text: Linguistic and Rhetorical Features*. Routledge.
- Hinkel, E. (Ed.), 1999. *Culture in Second Language Teaching and Learning*. Cambridge University Press.
- Hiver, P., Al-Hoorie, A.H., Vitta, J.P., Wu, J., 2021. Engagement in language learning: a systematic review of 20 years of research methods and definitions. *Language Teaching Res.* 1–30.
- Hyland, K., 1996. Talking to the academy: forms of hedging in science research articles. *Writ. Commun.* 13 (2), 251–281.
- Hyland, K., 1998a. Boosting, hedging and the negotiation of academic knowledge. *Text* 18, 349–382.
- Hyland, K., 1998b. *Hedging in Scientific Research Articles*. John Benjamins, Amsterdam, Netherlands.
- Hyland, K., 2000. *Disciplinary Discourses: Social Interactions in Academic Writing*. Longman, London.
- Hyland, K., 2001. Bringing in the reader: addressee features in academic articles. *Writ. Commun.* 18 (4), 549–574.
- Hyland, K., 2002. Authority and invisibility: authorial identity in academic writing. *J. Pragmat.* 34, 1091–1112.
- Hyland, K., 2004. *Disciplinary Discourses: Social Interactions in Academic Writing*. The University of Michigan Press, Ann Arbor.
- Hyland, K., 2005a. *Metadiscourse: Exploring Interactions in Writing*. Bloomsbury Publishing.
- Hyland, K., 2005b. Stance and engagement: a model of interaction in academic discourse. *Discourse Stud.* 7 (2), 173–192.
- Hyland, K., 2008. Disciplinary voices: interactions in research writing. *English Text Constr.* 1 (1), 5–22.
- Hyland, K., 2010. Metadiscourse: mapping interactions in academic writing. *Nordic J. Engl. Stud.* 9 (2), 125–143.
- Hyland, K., 2012. Undergraduate understandings: stance and voice in final year reports. In: Hyland, K., Guinda, C.S. (Eds.), *Stance And Voice in Written Academic Genres*. Palgrave Macmillan, London, pp. 134–151.
- Hyland, K., Milton, J., 1997. Qualification and certainty in L1 and L2 student writing. *J. Sec Lang. Writ.* 6, 183–205.
- Hyland, K., Tse, P., 2004. Metadiscourse in academic writing: a reappraisal. *Appl. Linguist.* 25 (2), 156–177.
- Ivanić, R., 1998. *Writing and Identity. The Discourse Construction of Identity in Academic Writing*. John Benjamins, Amsterdam.
- Ivanić, R., Camps, D., 2001. I am how I sound: voice as self-representation in L2 writing. *J. Sec Lang. Writ.* 10, 3–33.
- Juliaty, H., 2019. Exploring academic identities of EFL novice writers. *Indonesian J. Appl. Ling.* 9 (2), 324–334.
- Karahan, P., 2013. Self-mention in scientific articles written by Turkish and non-Turkish authors. *Procedia-Social Behav. Sci.* 70, 305–322.
- Kondowe, W., 2014. Hedging and boosting as interactional metadiscourse in literature doctoral dissertation abstracts. *Int. J. Lang. Learn. Appl. Ling. World* 5 (3), 214–221.
- Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33 (1), 159–174.
- Lee, J.J., Deakin, L., 2016. Interactions in L1 and L2 undergraduate student writing: interactional metadiscourse in successful and less-successful argumentative essays. *J. Sec Lang. Writ.* 33, 21–34.
- Lehman, I.M., 2018. *Authorial Presence in English Academic Texts: a Comparative Study of Student Writing across Cultures and Disciplines*. Peter Lang GmbH, Berlin.
- Lehman, I.M., Anderson, R., 2021. Inviting individual voice to second language academic writing. *Int. Rev. Pragmat.* 13 (1), 61–85.
- Liardet, C.L., 2018. 'As we all know': examining Chinese EFL learners' use of interpersonal grammatical metaphor in academic writing. *Engl. Specif. Purp.* 50, 64–80.
- Lorés-Sanz, R., 2011. The construction of the author's voice in academic writing: the interplay of cultural and disciplinary factors. *Text Talk* 31 (2), 173–193.

- Maley, A., Kiss, T., 2018. *Metadiscourse: Key to Developing Discourse Competence*. Routledge, New York, NY.
- Maton, K., 2007. Knowledge-knower structures in intellectual and educational fields. In: Christie, F., Martin, J.R. (Eds.), *Language, Knowledge and Pedagogy: Functional Linguistics and Sociological Perspectives*, pp. 87–108 (London, Continuum).
- Matsuda, P.K., 2001. Voice in Japanese written discourse: implications for second language writing. *J. Sec Lang. Writ.* 10, 35–53.
- Matsuda, P.K., Jeffery, J.V., 2012. Voice in student essays. In: *Stance and Voice in Written Academic Genres*. Palgrave Macmillan, London, pp. 151–165.
- Mauranen, A., 1993. *Cultural Differences in Academic Rhetoric*. Peter Lang.
- Myhill, D., Ahmed, A., Abdollazadeh, E., 2023. Going meta: bringing together an understanding of metadiscourse with students' metalinguistic understanding. *Lang. Teach.* 56 (1), 146–148.
- Ngampradit, K., 2020. A corpus-based study of metadiscourse boosters in applied linguistics dissertations written in Thailand and in the United States. In: Čermáková, Anna, Malá, Markéta (Eds.), *Variation in Time and Space: Observing the World through Corpora*. De Gruyter, Germany, pp. 321–350.
- 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., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372, 1–9.
- Petch-Tyson, S., 1998. Writer/reader visibility in EFL written discourse. In: Granger, S. (Ed.), *Learner English on Computer*. Longman, London and New York, pp. 107–118.
- Petrić, B., 2007. Rhetorical functions of citations in high- and low-rated master's theses. *J. Engl. Acad. Purp.* 6 (3), 238–253.
- Petrić, B., 2012. Legitimate textual borrowing: direct quotation in L2 student writing. *J. Sec Lang. Writ.* 21 (2), 102–117.
- Plonsky, L., Gass, S., 2011. Quantitative research methods, study quality, and outcomes: the case of interaction research. *Lang. Learn.* 61 (2), 325–366.
- Qin, W., Uccelli, P., 2019. Metadiscourse: variation across communicative contexts. *J. Pragmat.* 139, 22–39.
- Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., 1985. *A Comprehensive Grammar of the English Language*. Longman, New York.
- Reilly, J., Zamora, A., McGiverna, R.F., 2005. Acquiring perspective in English: the development of stance. *J. Pragmat.* 37, 185–208.
- Sun, Q., Kuzborska, I., Soden, B., 2022. Learning to construct authorial voice through citations: a longitudinal case study of L2 postgraduate novice writers. *System* 106, 102765.
- Swales, J., 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press, Cambridge/New York.
- Tang, R., John, S., 1999. The "I" in identity: exploring writer identity in student academic writing through the first person pronoun. *Engl. Specif. Purp.* 18, 23–39.
- Vande Kopple, W.J., 1985. Some exploratory discourse on metadiscourse. *Coll. Compos. Commun.* 36 (1), 82–93.
- Wei, X., Zhang, L.J., Zhang, W., 2020. Associations of L1-to-L2 rhetorical transfer with L2 writers' perception of L2 writing difficulty and L2 writing proficiency. *J. Engl. Acad. Purp.* 47, 100907.
- Wu, B., Paltridge, B., 2021. Stance expressions in academic writing: a corpus-based comparison of Chinese students' MA dissertations and PhD theses. *Lingua* 253, 103071.
- Wu, D., Buripakdi, A., 2021. Writer identity construction in EFL doctoral thesis writing. *GEMA Online J. Lang. Stud.* 21 (3), 16–36.
- Yasmin, T., Mahmood, M., 2017. A comparative analysis of writer's identity in Pakistani research theses of English and psychology. *Int. J. Engl. Ling.* 7 (6), 255–260.
- Yoon, H., 2017a. Textual voice elements and voice strength in EFL argumentative writing. *Assess. Writ.* 32, 72–84.
- Yoon, H., 2017b. Linguistic complexity in L2 writing revisited: issues of topic, proficiency, and construct multidimensionality. *System* 66, 130–141.
- Yoon, H.J., 2021. Interactions in EFL argumentative writing: effects of topic, L1 background, and L2 proficiency on interactional metadiscourse. *Read. Writ.* 34 (3), 705–725.
- Zhang, L., Zhang, L.J., 2021a. Development and Validation of the Questionnaire on EFL Students' Perceptions of Authorial Stance in Academic Writing. *Applied Linguistics Review*, 000010151520210003.
- Zhang, L., Zhang, L.J., 2021b. Fostering stance-taking as a sustainable goal in developing EFL students' academic writing skills: exploring the effects of explicit instruction on academic writing skills and stance deployment. *Sustainability* 13 (8), 4270.
- Zhao, C.G., 2019. Writer background and voice construction in L2 writing. *J. Engl. Acad. Purp.* 37, 117–126.