Journal of Language
Volume 8, Number 2, pp: 373-389
e-ISSN: 2685-8878 | p-ISSN: 2655-9080
DOI: https://doi.org/10.30743/jol.v8i2.14496

Foreign Language Speaking Anxiety Research in China (2005–2025): A Bibliometric and Content Analysis of CNKI Journal Articles

Yutong Hu, Weijie Hu
College of Foreign Languages, Zhejiang Normal University, Jinhua, China
E-mail: 360275875@qq.com

Article Info

Article history:
Received: October 15, 2026
Accepted: October 30, 2026
Published: November 10, 2026

Keywords:
bibliometric analysis;
CiteSpace;
content analysis;
foreign language teaching;
research hotspots;
speaking anxiety

ABSTRACT

Anxiety is a major affective factor in foreign language learning, particularly in speaking for Chinese learners with few authentic environments. This study maps how research on speaking anxiety in China has evolved, separating established findings from the problems that remain. The corpus is 262 articles on speaking anxiety among Chinese learners, published in Chinese-language journals indexed by CNKI, 2005–2025. A main search returned 299 records and a supplementary core-journal search added 13, giving 312 candidates; 50 were excluded at screening. It combines CiteSpace keyword analysis with article-level content coding on six dimensions. Three findings emerge. First, output follows a three-stage pattern — initial growth (62 articles, 2005–2010), rapid expansion (96, 2011–2015), stable fluctuation (104, 2016–2025); undergraduates account for 57.6% of the corpus. Second, the keyword network resolves into ten clusters, seven of which form two main lines: basic issues and pedagogical countermeasures. Third, burst keywords indicate a front moving from alleviating negative emotions towards positive-psychology constructs. The coding, however, reveals a documentation deficit. Of the 262 articles, 84.4% do not report their anxiety instrument's source, dimensions, reliability or validity and 80.2% declare no theoretical framework; 37.8% describe no data collection procedure, and only 6.5% report an experiment. These figures describe reporting practice in this corpus, not the field as a whole. The contribution is to show that increasing thematic consolidation has not been matched by methodological consolidation. Future research should strengthen theoretical grounding, standardise instrument reporting, and increase the use of longitudinal tracking and intervention designs.

Corresponding Author:
Yutong Hu
College of Foreign Languages, Zhejiang Normal University, Jinhua, China
Email: 360275875@qq.com

1. Introduction

Anxiety is a major affective factor in foreign language learning, showing a consistent moderate negative relationship with achievement (Horwitz, 2001), and it is especially consequential for speaking. Among skill-specific anxieties, speaking anxiety ranks second only to general foreign language classroom anxiety and exceeds reading, listening and writing anxiety (Y. Guo & J. Xu, 2014). Two features of the learning situation explain this. It requires learners to mobilise linguistic resources in real time, a process that language anxiety has been shown to disrupt (MacIntyre & Gardner, 1994). And classroom activities involving public performance are particularly likely to arouse it (Young, 1991). The Chinese context heightens these pressures. Authentic opportunities for oral practice outside the classroom are limited, and traditional cultural values and teacher evaluation shape the anxiety learners experience when speaking (Ma et al., 2009). The 2007 reform of college English teaching then made oral ability an explicit curricular goal (General Office of the Ministry of Education, 2007).

Two decades of research have accumulated, yet no review has mapped it using anything other than manual classification. Without such a map it is difficult to tell whether its growth has been accompanied by consolidation of its themes or of its methods, and the two questions call for different kinds of analysis. In this article, foreign language speaking anxiety refers to anxiety experienced in oral production and interaction. The construct is situation-specific, and Horwitz et al. (1986) located it in communication apprehension, one of the three related performance anxieties. It is treated here as a topic in its own right. Articles that measured speaking anxiety or a related oral variable alongside anxiety were included even when their stated construct was broader. The rules are given in Section 3.2.

The field has been reviewed before, but not in a way that answers the questions posed here. Reviews of foreign language anxiety have classified articles manually and reported frequency counts: J. Li and Q. Li (2016) analyse papers in foreign language journals over 2006–2015, and Shi and Xu (2013) analyse papers across 29 SSCI and 12 CSSCI journals over 1972–2011, both concluding that the range of research objects remains narrow; Shi and Xu add research methods, research perspectives, research settings and research topics to that list. Reviews of the speaking-anxiety subfield itself are fewer and thinner. Qu (2013) summarises work on the association between anxiety and oral learning, on the causes of speaking anxiety and on the proposed countermeasures. Gao (2023) reviews two decades along three dimensions, instruments, populations and research content, and judges the empirical base too thin, the populations too homogeneous and the methods too simple. None of these reviews uses keyword co-occurrence or clustering, and none codes reporting practice: whether an article names the instrument it administered, declares the theory it worked from, or describes the procedure by which its data were collected. The two reviews of the subfield state their conclusions about method as qualitative judgements, without reporting how many articles were affected. Conventional reviews therefore tell us what has been published, but not which themes hold the field together, when its research front shifted, or how well its methods have been reported.

The contribution of the study is twofold, and the two parts are not of equal weight. Its primary contribution is empirical. It provides a systematic map of the speaking-anxiety subfield in China that combines keyword-based bibliometric analysis with article-level content coding, showing both the thematic structure of the field and the research practices that produced it. Its secondary contribution is diagnostic. By coding instruments, frameworks and designs, the study quantifies weaknesses that earlier reviews have asserted but not measured. Neither route suffices alone. Bibliometric mapping cannot show what individual articles did, while content coding shows what they did but not how those practices relate to the field's thematic organisation.

The study answers three questions: (1) How is the publication output of research on speaking anxiety in China distributed over time, and what are the characteristics of its research populations? (2) What is the structure of the keyword co-occurrence network and its clusters, and which topics emerge as research hotspots? (3) How have the research themes evolved, and what problems remain in existing research with regard to measurement instruments, theoretical frameworks and research methods?

2. Literature Review

2.1 Previous Research on Foreign Language Speaking Anxiety in China

Research on foreign language speaking anxiety in China has grown steadily since 2005, but growth has not settled the questions it raises. First, there is agreement on what anxiety is but not on where it comes from. Studies examine individual differences such as self-efficacy (Y. Chen, 2019), sociocultural influences such as traditional cultural values and teacher evaluation (Ma et al., 2009), and classroom interaction and feedback, but these are ex-

Journal of Language
Volume 8, Number 2, pp: 373-389
e-ISSN: 2685-8878 | p-ISSN: 2655-9080
DOI: https://doi.org/10.30743/jol.v8i2.14496

amined separately, their relative weight is rarely estimated, and no integrated model appears to have been proposed. Second, the negative association between anxiety and speaking performance is well established, but its causal direction is not. Most of this evidence comes from general classroom anxiety rather than from speaking anxiety specifically. Cheng et al. (2007) and Luo (2019) both measured classroom anxiety with the FLCAS rather than speaking anxiety. Cross-sectional designs establish correlation only, and the one cross-lagged study identified in this corpus (M. Liu, 2011) has not been replicated. The one study in this corpus that tested for it directly found no facilitating anxiety at all (Luo, 2019), so the facilitating side of the construct is even less settled. Third, intervention research is the most numerous category (73 of the 262 articles, or 27.9%) and the weakest in evidential terms. The methods tested — rational emotive behaviour therapy (T. Wang, 2014), pronunciation and intonation imitation training (Wu & Zhao, 2013), cooperative learning (Xiao, 2025) — are diverse but rarely replicated and seldom linked to instrument validity, so no approach is yet independently supported. All three methods report positive effects, which is itself a reason for caution. Fourth, the turn towards positive psychology is recent and narrow, currently centred on enjoyment and willingness to communicate (Fan & Li, 2024; Y. Wang, 2023).

2.2 Research Gap

These four observations describe what the field has found. What its reviewers have not done is relate those findings to the practices that produced them. The three reviews that cover both the themes and the practices behind them (J. Li & Q. Li, 2016; Shi & Xu, 2013; Gao, 2023) describe the two separately and never relate them. It therefore remains unknown whether the consolidation visible in the keyword network (Section 4.2) has been matched by consolidation in how research is designed and reported (Section 4.4), which is the question this study addresses in Section 4.5.

2.3 Conceptual and Analytical Framework

This study concerns foreign language speaking anxiety, whose investigation in China originates in Horwitz et al.'s (1986) account of foreign language classroom anxiety as a distinct complex of self-perceptions, beliefs, feelings and behaviours related to classroom language learning. In that account, classroom anxiety is theorised in relation to communication apprehension, test anxiety and fear of negative evaluation, and speaking anxiety is treated here as the narrower construct that comes closest to the first of them. These three anxieties supply the background against which Chinese work is read. They are not coded as separate categories. The study does not test this construct; it maps how researchers in China have approached it. Two analytical approaches guide that mapping, and neither is a theory of speaking anxiety: bibliometrics, in which keyword co-occurrence and clustering are taken to reveal a field's intellectual structure, and content analysis, in which each article is classified by what it explicitly reports. The anxiety instrument is coded separately from the research design, since administering a scale says nothing about a study's logic of inquiry. The bibliometric component works on article keywords, so the corpus is topically bounded before coding begins. Content analysis, by contrast, treats each article as a report of what its authors did. That is why the six dimensions record reported rather than observed practice. Theoretical framework is coded for the same reason: whether a framework is declared bears directly on whether the field's findings are cumulative. Operational definitions and decision rules are given in Section 3.2 for the dimensions whose boundaries could not be inferred from the category labels, and the remaining categories are named in ordinary language in Table 2.

3. Research Method

3.1 Data Sources and Search Strategy

This study used the CNKI academic journal database as its data source. The search covered the period from 1 January 2005 to 31 December 2025 and was limited to journal articles. Both searches were run on 5 October 2026 through the CNKI institutional interface, restricted by the document-type filter to journal articles and by language to Chinese; theses, conference papers and newspapers were excluded at this stage. CNKI exposes no version number to users. CNKI was chosen as the most comprehensive index of Chinese-language journals in language education; the corpus therefore excludes English-language publications, theses, conference papers and studies of non-Chinese learners, so the findings describe the Chinese-language journal literature rather than the international field. Chinese-language literature was retrieved using the professional search function, with SU as the subject field, which covers titles, keywords and abstracts. The "Chinese–English expansion" option was enabled, whereas "synonym expansion" was disabled. The main search string was as follows (given in Chinese, as CNKI requires search terms to be entered verbatim):

SU=('口语焦虑'+'口语交际焦虑'+'口语表达焦虑'+'外语焦虑'+'英语焦虑'+'语言焦虑'+'课堂焦虑'+'交际焦虑'+'交流焦虑'+'交际畏惧'+'交流畏惧') AND SU=('口语'+'口试'+'口头'+'口语产出'+'口语表达'+'口语水平'+'口语能力'+'口语交际'+'交际意愿'+'说英语')

(In English, SU=('speaking anxiety'+'oral communicative anxiety'+'oral expression anxiety'+'foreign language anxiety'+'English anxiety'+'language anxiety'+'classroom anxiety'+'communicative anxiety'+'communication anxiety'+'communication apprehension'+'communication fear') AND SU=('speaking'+'oral examination'+'oral'+'oral output'+'oral expression'+'oral proficiency'+'oral ability'+'oral communication'+'willingness to communicate'+'speaking English'). CNKI requires the Chinese terms; this rendering is given so that readers without Chinese can follow the logic of the string, and it corresponds term for term to the Chinese above.)

This string required articles to address both anxiety and speaking, and returned 299 records. A single round of searching may have missed some articles, particularly those that investigate speaking yet use only broader terms such as "foreign language anxiety" or "classroom anxiety" in their titles, keywords and abstracts, which the above string would have excluded. A supplementary search was therefore conducted, limited to Peking University Core journals and CSSCI-indexed journals, in which the second group of terms was broadened to superordinate terms such as "foreign language", "English", "second language" and "L2". The search string was:

SU=('口语焦虑'+'口语交际焦虑'+'口语表达焦虑'+'外语焦虑'+'英语焦虑'+'语言焦虑'+'课堂焦虑'+'交际焦虑') AND SU=('外语'+'英语'+'二语'+'第二语言'+'口语'+'口试'+'口语产出')

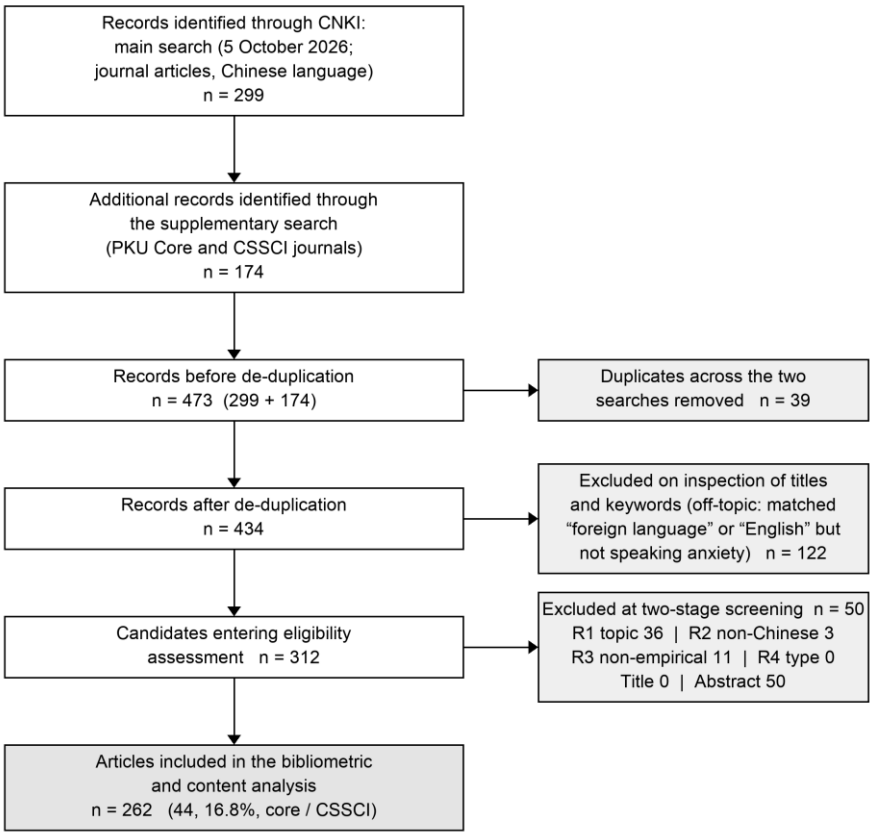
(In English, SU=('speaking anxiety'+'oral communicative anxiety'+'oral expression anxiety'+'foreign language anxiety'+'English anxiety'+'language anxiety'+'classroom anxiety'+'communicative anxiety') AND SU=('foreign language'+'English'+'second language'+'L2'+'speaking'+'oral examination'+'oral output'))

This string returned 174 records. A record-by-record comparison with the main search showed 39 overlaps, leaving 135 non-duplicate records, of which 122 were off-topic on inspection of titles and keywords and 13 were eligible and new (Figure 1). The two rounds therefore yielded 299 + 13 = 312 candidate records. After two-stage screening by title and abstract, 50 records were excluded, leaving 262 articles. Screening was carried out by the first author. Cases whose abstract was ambiguous were resolved by consulting the full text where it was available, and a record was retained whenever its abstract reported both an anxiety measure and a speaking outcome, even if neither was the article's stated focus. The screening is reported in the sequence set out in the PRISMA 2020 statement (Page et al., 2021). The restriction was deliberate: the supplementary round was a check on the main search rather than an extension of it. Because its 13 additional records were merged into the same corpus and

Journal of Language
Volume 8, Number 2, pp: 373-389
e-ISSN: 2685-8878 | p-ISSN: 2655-9080
DOI: https://doi.org/10.30743/jol.v8i2.14496

coded identically, it affects the composition of the corpus rather than the coding; the proportion of core-journal articles is reported in Section 3.2 so that readers can judge the effect. The residual risk of missing an article that discusses speaking anxiety without using any term from either string was tested by the supplementary round, whose superordinate terms add only 13 new eligible records to the 299 of the main search. Unlabelled studies cannot be ruled out by searching.

Figure 1
Screening and Inclusion Flow of the Corpus



Note. N = 262. Both searches required one anxiety term and one speaking term in the subject field (SU); the supplementary round broadened the second group to superordinate terms.

3.2 Literature Screening and Coding

Two inclusion criteria had to be met. The first concerned the speaking dimension: an article qualified if its research questions, variables or content involved oral expression, oral production, speaking fluency or proficiency, speaking scores, willingness to communicate, pronunciation and intonation, or speaking English. The second concerned the anxiety dimension: the article had to involve anxiety-related concepts such as speaking anxiety, foreign language anxiety, classroom anxiety, communication anxiety or communication apprehension.

Four exclusion criteria were defined and applied in turn. Topic mismatch (R1) refers to articles whose content concerns neither speaking nor anxiety, and it proved the most common category. Articles whose research population was not Chinese learners were coded as R2. Articles consisting solely of reviews, commentaries, research updates or meta-analyses were coded as R3 and excluded (n = 11), nearly all of them literature reviews of foreign language anxiety or speaking anxiety. Articles reporting the authors' own teaching experience,

classroom observations or reflective accounts were retained, because they are first-hand accounts rather than syntheses. The two cannot be told apart from titles and abstracts alone, which is why such papers are described in Section 4.4 as experience-based or discursive. Book reviews, editorials and conference reviews were coded as R4 for not matching the required publication type. Topical relevance, research population, empirical character and publication type were checked in turn, and an article was excluded as soon as it failed any one of these checks.

Screening proceeded in two stages. Titles and keywords were examined first. No record was excluded at this stage, because both search strings already required one anxiety term and one speaking term. Abstracts were then read to confirm that the research questions or variables did in fact involve the speaking dimension; articles that did not were excluded. The two-stage screening removed 50 records in total: 36 for topic mismatch (R1), 11 as non-empirical studies (R3) and 3 for non-Chinese learner samples (R2), with no records excluded for publication type (R4). This left 262 articles for final inclusion, of which 44 (16.8%) appeared in Peking University Core journals or CSSCI-indexed journals.

All 262 articles were then coded along six dimensions (following the coding logic set out by Krippendorff, 2018): research population, research design, data collection method, anxiety instrument, theoretical framework and research focus. The categories used for each dimension are set out in Table 2 (Section 4.4). Where a study drew on more than one learner population, it was coded by the population that supplied the majority of the sample; studies reporting balanced samples from more than one level were coded as mixed (n = 1). Coding was carried out article by article in Excel.

Some labels carry different meanings in different dimensions. “Not reported” means that the relevant information could not be found in the article, and the two dimensions where this is least self-evident are stated explicitly below. Within each dimension the categories are mutually exclusive, so each column in Table 2 sums to 262. Across dimensions the codes are independent, though in practice they are associated: quantitative studies used only questionnaires or experiments, and qualitative studies only interviews or no stated procedure. Experiment is coded as a data-collection category because it names the procedure used, whereas design records the mode of reasoning. An experiment is therefore quantitative in design and experimental in data collection, which is why the two are reported separately.

Two dimensions require an explicit rule, both fixed before coding began. Anxiety instrument: an article is coded as reporting an instrument only if the instrument’s source is identifiable from the article itself — naming the FLCAS is sufficient, naming nothing is not, and an article that evidently administered a scale without identifying it is coded as not reporting, so this code records reporting practice rather than whether an instrument was used. Theoretical framework: an article is coded as having a framework only if a theory is named or unambiguously identifiable as the organising lens of the study; theory cited in passing, without organising the design or the variables, is coded as not reported.

All references cited in this article were checked against the original publications, and every in-text citation corresponds to an entry in the reference list; no identifier was inserted without verification.

3.3 Analytical Tool and Parameter Settings

Keyword co-occurrence and clustering analyses were performed with CiteSpace 6.4.R1 (64-bit) Advanced (C. Chen, 2006), using Keyword as the node type, a time slice spanning 2005 to 2025 (slice length = 1 year), and the g-index (k = 25) as the selection criterion. The remaining parameters (link strength, look-back years and pruning method) are listed in Table 1.

Journal of Language
Volume 8, Number 2, pp: 373-389
e-ISSN: 2685-8878 | p-ISSN: 2655-9080
DOI: https://doi.org/10.30743/jol.v8i2.14496

Three parameter choices warrant comment. The g-index (k = 25) was used rather than a fixed top-N threshold because it allows the number of selected nodes to scale with each slice's profile while keeping the network at a size that can be labelled and read. Pathfinder pruning with "pruning sliced networks" was applied because it retains structurally essential links while removing dense local connections that would otherwise obscure cluster boundaries. The burst-detection parameters (γ = 0.5, minimum duration = 2) follow the software's recommended defaults for corpora of this size. Because clustering outcomes are sensitive to keyword selection, terminology standardisation and parameter settings, no systematic sensitivity analysis was conducted; the parameters were set before the analysis and not varied afterwards. Keyword-based analysis has three limits that this paper observes. It records the terms authors chose to index an article rather than the content of the article, so a study may use "speaking anxiety" in its text without listing it as a keyword. It treats different labels for the same construct as separate nodes unless they are standardised, and standardisation is partly a matter of judgement. And it measures co-mention rather than intellectual dependence, because CNKI records carry no reference field; co-citation and bibliographic coupling analyses, which would capture dependence directly, could not therefore be performed. The findings reported below are therefore keyword-based thematic relationships, not a reconstruction of the field's intellectual foundations.

Table 1
Parameter Settings and Structural Indicators of the Keyword Co-occurrence Network

Category	Indicator	Value / Setting
Parameter settings	Time slicing	2005–2025 (Slice Length = 1)
	Node type	Keyword
	Selection criteria	g-index (k = 25)
	Link parameters	LRF 2.5; L/N 10; LBY 5; e 1.0
	Link strength and scope	Cosine; Within Slices
	Pruning	Pathfinder + Pruning sliced networks
Structural indicators	Number of nodes (*N*)	253
	Number of links (E)	541
	Network density	0.017
	Largest connected component	222 (87%)
	Modularity (*Q*)	0.619
	Weighted mean silhouette (*S*)	0.868
	Number of clusters	10

Note. LRF = link retention factor; L/N = maximum links per node; LBY = look-back years; e = selection threshold. By convention, Q > 0.3 indicates a significant clustering structure, and S > 0.7 indicates credible clustering results.

4. Discussion

4.1 Temporal Distribution of Publication Output and Research Populations

Between 2005 and 2025, 262 journal articles were published on foreign language speaking anxiety in China, averaging approximately 12.5 articles per year. As Figure 2 shows, publication output exhibits a three-stage pattern of initial growth, rapid expansion and stable fluctuation. The three periods were defined by changes in the annual counts rather than imposed in advance.

Output rose from three articles in 2005 to 14 in 2010. These six years, the initial growth stage, produced 62 articles, or 23.7% of all articles. Work in this period centred on in-

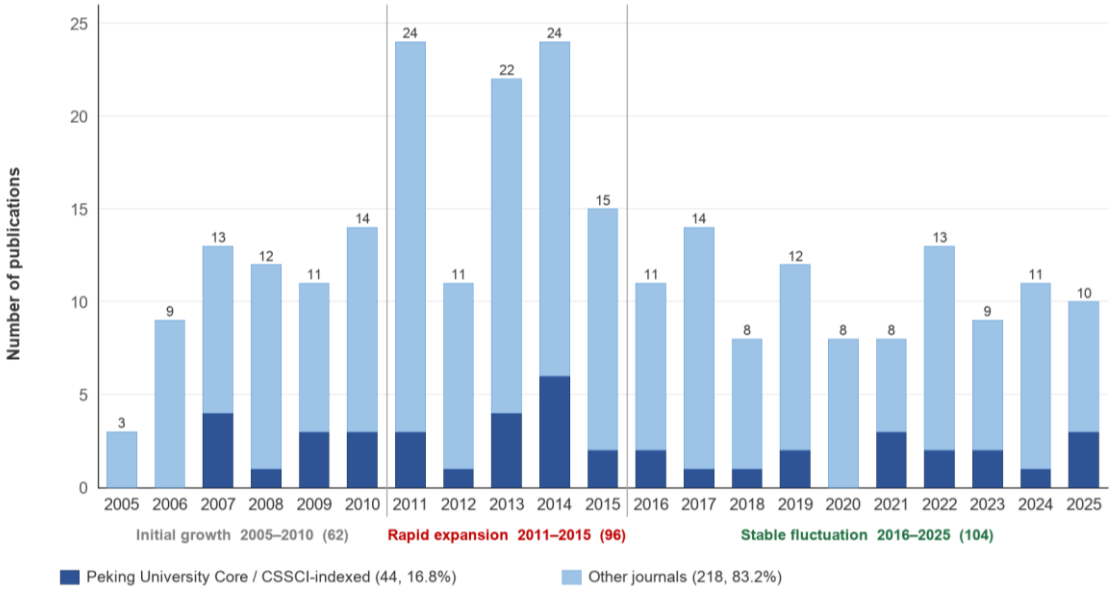
troducing classroom anxiety and validating anxiety scales; speaking anxiety was still a dimension mentioned in passing rather than an independent topic.

Publication output rose markedly between 2011 and 2015, with 96 articles over the five years (36.6%) and peaks of 24 in both 2011 and 2014. This may plausibly be associated with the emphasis placed on speaking ability by the reform of college English teaching, though no evidence in the present data establishes the association causally. The 2007 revision of the College English Curriculum Requirements called for a comprehensive ability to use English, especially in listening and speaking (General Office of the Ministry of Education, 2007). Articles on "sources of anxiety" increased from 11 in the initial growth stage to 30, their share rising from 17.7% to 31.3%, which suggests that the second stage added a concern with origins to the existing concern with remedies.

The period 2016–2025 was one of stable fluctuation, with 104 articles (39.7%) and an average of about 10.4 per year, ranging narrowly between 8 and 14 with no clear trend. These figures are consistent with increasing thematic consolidation. However, the number of core-journal articles did not rise in parallel: after 2016 the annual figure was mostly 1–3 and fell to zero in 2020. This may indicate that high-quality journals have not become more receptive to the topic.

Figure 2

Annual Distribution of Publication Output on Foreign Language Speaking Anxiety in China, 2005–2025



Note. Dark bars represent articles published in Peking University Core or CSSCI-indexed journals; light bars represent articles published in other journals.

In terms of research population (Table 2), undergraduates were the most frequently studied group, appearing in 151 articles (57.6%), followed by higher vocational college students (37, 14.1%), junior secondary students (7, 2.7%), postgraduate learners (4, 1.5%), senior secondary students (3, 1.1%) and mixed-level samples (1, 0.4%). A further 59 articles (22.5%) did not report the level of their research population, usually describing it only in general terms such as "college students", "English learners" or "Chinese students". Research has thus concentrated on undergraduates and, secondarily, on higher vocational college students, leaving basic education and postgraduate learners under-studied.

Journal of Language
Volume 8, Number 2, pp: 373-389
e-ISSN: 2685-8878 | p-ISSN: 2655-9080
DOI: https://doi.org/10.30743/jol.v8i2.14496

4.2 Keyword Co-occurrence and Clustering

CiteSpace produced a network of 253 nodes and 541 links (density 0.017); the largest connected component contains 222 nodes (87%). Modularity Q is 0.619, above the empirical threshold of 0.3, and weighted mean silhouette S is 0.868, above the credibility threshold of 0.7, indicating a significant clustering structure with high within-cluster homogeneity.

"Speaking anxiety" ranked first on both counts, with 58 occurrences and a centrality of 0.60. Three indicators are reported and should not be conflated: occurrence frequency is the number of articles in which a keyword appears; centrality is a betweenness-based measure of how often a node lies on the shortest path between other nodes; and cluster membership reflects co-occurrence with a group of terms. Its centrality is more than twice that of the next retained term, "English speaking" (23 occurrences, 0.29), suggesting that speaking anxiety occupies a pivotal position linking the various thematic areas. Centrality is a structural property of the network, not a measure of research quality or theoretical importance: it shows that a term is widely co-mentioned, not that the construct is better developed. "Foreign language anxiety" (20 occurrences, 0.20) and "language anxiety" (15, 0.12) form a layer of superordinate concepts, reflecting a lack of full terminological consensus. Generic and superordinate terms were excluded from the frequency counts; "anxiety" itself had a raw frequency of 40 and a centrality of 0.42, ranking second on both.

"Willingness to communicate" (13 occurrences, 0.09) and "communication anxiety" (8, 0.09) show that research has moved beyond measuring anxiety towards its relationship with communicative orientation. "Speaking instruction" (11), "classroom anxiety" (10) and "oral expression" (7) reflect the classroom context, while "college students" (9, 0.08) and "higher vocational college students" (6, 0.10) mark the research population.

The cluster analysis produced 10 clusters, ranging in size from 11 to 39 keywords, with silhouette values between 0.762 and 0.933. Apart from cluster #3 (0.762), which is slightly lower, all other values exceed 0.8, indicating good clustering quality.

In terms of their content, seven of the ten clusters can be grouped into two main lines, with the remaining three supplementing them. Clusters were grouped on the basis of three criteria applied to their principal keywords and representative articles: shared research question (what the cluster asks), shared variable type (what it measures) and shared methodological orientation. Clusters that ask how anxiety is distributed and measured form the first line; those that ask how it can be reduced form the second. The three remaining clusters were judged supplementary because their keywords cut across both lines rather than constituting a third research programme. The larger line concerns basic issues and comprises clusters #0, #6 and #7, with 78 keywords in total: survey studies of speaking anxiety and individual differences (#0, 39 keywords, average year 2014), the relationship between classroom anxiety and oral expression (#6, 20, 2011) and empirical research and instrument development (#7, 19, 2013). Their common ground lies in measuring anxiety levels and examining related variables such as gender, achievement and self-confidence; representative studies include Cheng et al. (2007) and Y. Guo and J. Xu (2014).

The other concerns pedagogical countermeasures and comprises clusters #3, #5, #8 and #9, with 70 keywords in total. Its core terms are "countermeasures", "coping strategies" and "teaching strategies", and it addresses how pedagogical interventions can reduce speaking anxiety. The methods involved include cooperative learning and task-based teaching; representative studies include J. Wang and B. Wang (2014) and P. Li (2014).

Clusters #1, #4 and #2 supplement these main lines from three perspectives: affective factors, the teaching context and the integration of multiple emotions. Cluster #1 concerns affective factors and the causes of speaking anxiety (27 keywords, average year 2010), while cluster #4 concerns the psychology of anxiety in speaking instruction (23, 2012). Cluster #2

has the most recent average year (2016), and its keywords cover foreign language anxiety, willingness to communicate, foreign language enjoyment, communication apprehension and emotional intelligence, suggesting a move from a single anxiety variable towards the integration of multiple emotional variables.

4.3 Keyword Bursts and Thematic Evolution

To examine the temporal evolution of research themes, burst detection was performed on the keywords with the parameters $\gamma = 0.5$ and Minimum Duration = 2, identifying 11 burst keywords (Figure 3). By burst start year, the thematic evolution of research on speaking anxiety in China can be divided into three stages. The boundaries were derived from the burst start years themselves rather than from fixed calendar periods: the stages contain bursts beginning in 2007–2011, 2017–2018 and 2020–2023 respectively. A keyword was assigned to a stage by its burst start year, not by the year the topic first appeared; the stages are therefore groupings of the burst chronology, not periods of equal length.

The first stage contains six burst keywords and centres on basic issues and pedagogical countermeasures: “speaking instruction”, “classroom anxiety” and “English classroom” for early work on classroom anxiety; “countermeasures” and “strategies” for the quick turn to remedies (J. Guo & Liu, 2007); and “empirical research” from 2011 for the shift away from experience-based papers, visible in the corpus itself, since the share of articles reporting a data-collection method rises from 59.7% before 2011 to 63.0% afterwards. The second stage contains two keywords. “Foreign language anxiety” has the highest burst strength of all (3.68), from 2017 to the present; “coping strategies” burst from 2018 to 2022, indicating continued interest in practical responses. The first of these bursts shows that the perspective expanded from specific contexts to the overarching construct and to outcome variables such as speaking participation (Luo, 2019) and overall English proficiency (Fei, 2016).

The third stage contains three keywords and suggests an emerging tendency towards positive psychology, also observed internationally (Dewaele & MacIntyre, 2014). “Willingness to communicate” has been bursting since 2020 and “foreign language enjoyment” appeared only in 2023; together they suggest a shift from alleviating negative emotions towards cultivating positive ones. Two keywords are thin evidence for a field-wide transformation, however, the shift is an emerging tendency rather than a completed one. “Teaching strategies” burst from 2020 to 2023, so the countermeasure orientation had not faded by the end of the period. Methods are also more refined. Gu et al. (2025), for example, combined real-time learner self-assessment with speech-function analysis to track changes in willingness to communicate.

Figure 3

Keyword Burst Map

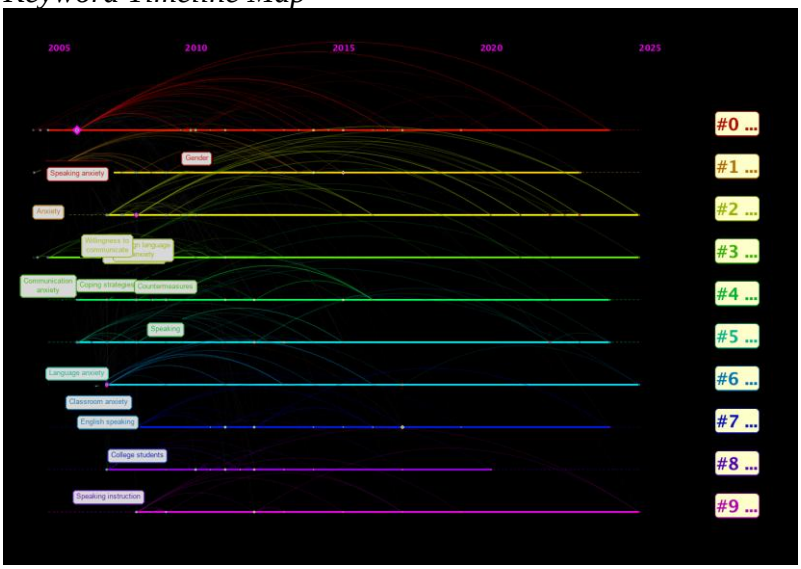


Note. Red bars indicate the burst period of each keyword, and bar length corresponds to the duration of the burst.

Journal of Language
Volume 8, Number 2, pp: 373-389
e-ISSN: 2685-8878 | p-ISSN: 2655-9080
DOI: https://doi.org/10.30743/jol.v8i2.14496

The timeline map (Figure 4) is consistent with this assessment. "Speaking anxiety" appeared as early as 2006 and runs throughout, "classroom anxiety" and "speaking instruction" emerged around 2007, and "foreign language enjoyment" is the most recent among the higher-frequency terms. After 2018 the number of newly added keywords declined markedly. This may indicate stabilisation and deepening, but three alternatives cannot be excluded: reporting practice may have standardised, database coverage may have changed, and terminology may have converged without any change in what is studied.

Figure 4
Keyword Timeline Map



Note. The horizontal axis shows the year and the vertical axis the cluster number; node position indicates the year in which a keyword first appeared, and node size indicates frequency. Clusters: #0 survey studies of speaking anxiety and individual differences; #1 affective factors and the causes of speaking anxiety; #2 foreign language anxiety, willingness to communicate and foreign language enjoyment; #3 causes of communication anxiety and coping strategies; #4 anxiety in college English speaking instruction; #5 language anxiety and teaching strategies for speaking ability; #6 classroom anxiety and oral expression; #7 empirical research and instrument development; #8 anxiety in English for specific purposes instruction; #9 anxiety profiles and cooperative learning interventions among higher vocational college students. Clusters #0, #6 and #7 form the basic-issues line; #3, #5, #8 and #9 the pedagogical-countermeasure line; #1, #4 and #2 supplement these two lines (Section 4.2).

4.4 Research Methods, Theoretical Frameworks and Research Focus: Distribution and Reflections

The coding results for the 262 articles (Table 2) reveal several noteworthy features. Research designs are relatively balanced: 103 articles (39.3%) were quantitative, 100 (38.2%) qualitative and 59 (22.5%) mixed-methods. Data collection, however, was markedly uneven: 99 articles (37.8%) were experience-based or discursive and reported no explicit data collection procedure. Close reading shows that 97 of the 99 present the authors' own teaching experience, classroom observations or reflective reasoning, and only two report survey or interview evidence too thinly described to be coded as systematic. These papers are therefore distinct from the reviews excluded at screening (R3, n = 11), which consisted solely of syntheses of other researchers' work. A further 78 (29.8%) used questionnaires and 67 (25.6%) combined questionnaires with interviews. Only 17 (6.5%) reported an experiment; existing inter-

vention studies, such as those by T. Wang (2014), J. Wang and B. Wang (2014) and P. Li (2014), were reasonably rigorous, but their number is too small relative to the rest of the corpus. The corpus is therefore predominantly descriptive or correlational rather than experimental, and as a whole it can support claims about association but not about causal direction.

Table 2

Coding Distribution of the 262 Articles

Coding dimension	Value	n	%
Research population	Undergraduate	151	57.6
	Not reported	59	22.5
	Higher vocational college	37	14.1
	Junior secondary	7	2.7
	Postgraduate	4	1.5
	Senior secondary	3	1.1
	Mixed	1	0.4
Research design	Quantitative	103	39.3
	Qualitative	100	38.2
	Mixed-methods	59	22.5
Data collection method	Other (experience-based)	99	37.8
	Questionnaire	78	29.8
	Mixed	67	25.6
	Experiment	17	6.5
	Interview	1	0.4
Anxiety instrument	Not reported	221	84.4
	FLCAS	17	6.5
	Other existing scales	14	5.3
	Speaking-anxiety-specific scales	7	2.7
	Self-designed scales	3	1.1
Theoretical framework	Not reported	210	80.2
	Foreign language anxiety theory	27	10.3
	Willingness to communicate (WTC) theory	20	7.6
	Complex dynamic systems theory	3	1.1
	Positive psychology	2	0.8
Research focus	Intervention effects	73	27.9
	Sources of anxiety	59	22.5
	Influencing factors	54	20.6
	Relationship with speaking performance	39	14.9
	Anxiety status	37	14.1

Note. N = 262. "Not reported" indicates that the article did not provide the relevant information; "other" under data collection method refers to experience-based and discursive articles with no explicit data collection method.

Instrument reporting and theoretical grounding were both weak. 221 articles (84.4%) did not report the source, dimensions or measurement properties of the scale used. This code records reporting practice rather than the absence of an instrument. An article that identified the scale but not its properties would have been coded the same way. The finding documents

Journal of Language
Volume 8, Number 2, pp: 373-389
e-ISSN: 2685-8878 | p-ISSN: 2655-9080
DOI: https://doi.org/10.30743/jol.v8i2.14496

a documentation deficit, not invalid instruments. These elements are not equally applicable, since source and dimensionality apply to every instrument whereas reliability and validity apply only where a scale was named. The figure therefore combines unnamed instruments with named ones whose properties go unreported, and separating the two would need full-text coding. Among the remaining 41 articles, only 17 (6.5%) explicitly used the FLCAS, 14 (5.3%) adopted other existing scales, 7 (2.7%) developed speaking-anxiety-specific scales and 3 (1.1%) relied on self-designed scales. Because the scales differ, reported means and correlation coefficients lack an interpretable baseline, and reporting is weakest in the countermeasure-oriented clusters and most explicit in cluster #2. Similarly, 210 articles (80.2%) declared no theoretical framework; among those that did, foreign language anxiety theory ranked first (27, 10.3%), followed by willingness to communicate theory (MacIntyre et al., 1998) (20, 7.6%), complex dynamic systems theory (3, 1.1%) and positive psychology (2, 0.8%). The figure records explicit declaration rather than theoretical reasoning as such, since a study may draw on a framework without naming it; what it establishes is that the framework is not visible to a reader, not that none informed the work. Most studies therefore present variables without stating the basis on which they were selected, though a few in cluster #2 are moving towards the L2 motivational self system and complex dynamic systems theory (Y. Chen, 2019; Yu & Liu, 2021).

With respect to research focus, intervention effects were the most frequently studied topic (73 articles, 27.9%), followed by sources of anxiety (59, 22.5%), influencing factors (54, 20.6%), the relationship with speaking performance (39, 14.9%) and anxiety status (37, 14.1%). The five categories were distinguished by asking what each article set out to explain: whether it evaluated an intervention, accounted for causes, tested correlates, estimated association with performance, or described levels. The first two together account for more than half of the corpus, so the agenda is weighted towards origins and remedies, while process-oriented questions receive little attention. This matters for the third research question, since the two largest categories are also the two that depend most on the instrument and framework work shown above to be thin.

4.5 Research Landscape and the Shift of the Research Front

Read together, the analyses answer the three research questions consistently, and the content analysis supplies the methodological counterpart to the bibliometric picture, since instrument and framework reporting is weak and experimental procedures remain rare. Two features of the resulting landscape deserve comment.

First, the network concentrates around “speaking anxiety”, whose centrality (0.60) is twice that of the next retained term. Concentration makes a field recognisable, but a network in which most studies are variations on a single theme offers little evidence of theoretical innovation. Second, the two main lines are comparable in size (78 versus 70 keywords) but weakly connected: instruments and individual-difference findings rarely appear in intervention designs, and proposed countermeasures rarely return to basic research for mechanism testing. The two lines are analytical groupings rather than disjoint sets, so weak connection means the literatures seldom cite each other, not that they are mutually exclusive. This is an interpretation of network structure rather than a demonstrated absence of knowledge transfer; testing it directly would require citation or full-text linkage, which the CNKI records do not support. Three interventions cited in Section 2.1 illustrate the pattern. T. Wang (2014) applies rational emotive behaviour therapy without engaging Luo’s (2019) finding that facilitating anxiety does not exist; Wu and Zhao (2013) train pronunciation without reporting an instrument source; Xiao (2025) tests cooperative learning without declaring a framework. Each is coherent, and none builds on the work it presupposes.

What the analysis does establish is that thematic consolidation has not been matched by methodological consolidation. The stability visible in the keyword network is therefore consistent with a field that has settled on its topics without settling on its methods. An integrated design would take a finding from the basic line, such as the absence of facilitating anxiety or the moderating role of self-efficacy, and test it inside an intervention, measuring both the outcome and the mechanism with a reported instrument. None of the 73 intervention studies in this corpus does both.

5. Conclusion

The three research questions can now be answered directly. First, output follows a three-stage pattern — initial growth (2005–2010, 62 articles), rapid expansion (2011–2015, 96) and stable fluctuation (2016–2025, 104) — concentrated on undergraduates (151, 57.6%), with basic education and postgraduate learners under-represented. Second, the network centres on "speaking anxiety" (58 occurrences, centrality 0.60), and the ten clusters resolve into two substantive lines, basic issues (78 keywords) and pedagogical countermeasures (70), with three supplementary clusters. Third, the research front has moved from classroom description and countermeasure proposals (2007–2011), through consolidation of the overarching construct (2017), to positive-psychology constructs (2020–2023); yet the methodological weaknesses identified in Section 4.4 have not been addressed in step with this thematic development. The three are connected rather than parallel. Consolidation at the thematic level was achieved by accumulating studies that share vocabulary but not instruments, theoretical frameworks or designs, which is why the field can look integrated in a co-word map while remaining incomparable at the level of findings. The contribution of the study is to show that the two analytical routes converge on this diagnosis: thematically the field has consolidated, methodologically it has not.

Three limitations should be kept in view. The corpus is restricted to Chinese-language articles indexed in CNKI, and because those records contain no reference field only keyword-based analysis was possible. The findings therefore describe this corpus rather than international research, and the intellectual base of the field remains uncharacterised.

Term choice involves judgement at two points, in selecting the search strings and in normalising the keywords, and a differently worded set of terms would have returned a partly different corpus.

Screening and coding were carried out by a single researcher. This is defensible for screening, since both search strings already fixed the topic and the abstract stage turned on whether a speaking variable was present. It is more exposed for coding, but the rules were fixed in advance and are stated in full, so a reader can re-code any article and check the result. The instrument, framework and population percentages, whose rules leave least room for judgement, are the most secure of the coded findings, and the bibliometric results do not depend on coding at all. Where judgement could enter it would have entered in one direction: a coder willing to infer a framework from a passing citation would have produced higher counts, and likewise for instruments identifiable only from a scale description. The data-collection figure is the one that rests on a genuinely indeterminate boundary, and it should be read as indicative rather than exact. Two qualifications follow. The dataset records what has been published, not what has been achieved in classrooms, and a rising keyword shows attention rather than changed practice.

Four recommendations follow from these findings, in descending order of urgency. First, instrument reporting should become a condition of publication, since without a common baseline results cannot be compared; the 84.4% recorded in Section 4.4 is the direct consequence of its absence. Second, theoretical frameworks should be declared explicitly, so that variables are justified rather than merely assembled; with 80.2% declaring none, that justifica-

Journal of Language
Volume 8, Number 2, pp: 373-389
e-ISSN: 2685-8878 | p-ISSN: 2655-9080
DOI: https://doi.org/10.30743/jol.v8i2.14496

tion is currently missing. Third, designs should be strengthened. Cross-lagged, longitudinal and quasi-experimental studies can address causal direction and intervention effects, which the present corpus rarely supports, since only 6.5% report an experiment. Fourth, research populations should be extended beyond undergraduates and outcome measures broadened from reduced anxiety to willingness to communicate and enjoyment; international comparison using databases such as Web of Science would offset the single-database limitation acknowledged above.

Funding

This work is supported by the National Social Science Fund of China (NSSFC) Project "Construction of an Automated Speaking Assessment Model for International Chinese Language Education" (Grant No.: 25BYY143).

References

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, *57*(3), 359–377. https://doi.org/10.1002/asi.20317

Chen, Y. (2019). A study of the relationships among the L2 motivational self system, self-efficacy, language anxiety and motivated classroom oral participation behaviour [in Chinese]. *Foreign Language Research*, (1), 63–70. https://doi.org/10.16263/j.cnki.23-1071/h.2019.01.011

Cheng, Y., He, Q., & Han, X. (2007). The effects of English classroom anxiety on college students' oral expression [in Chinese]. *Language Teaching and Linguistic Studies*, (1), 89–95.

Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching*, *4*(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5

Fan, Y., & Li, L. (2024). The influence of emotional intelligence on L2 willingness to communicate: The mediating role of academic emotions [in Chinese]. *Modern Foreign Languages*, *47*(5), 653–664. https://doi.org/10.20071/j.cnki.xdwy.20240724.010

Fei, W. (2016). An empirical study of the effects of foreign language learning anxiety on college students' comprehensive English application ability in a multimodal online environment [in Chinese]. *Technology Enhanced Foreign Language Education*, (3), 68–73+79. https://doi.org/10.20139/j.issn.1001-5795.2016.03.012

Gao, L. (2023). A review of research on foreign language speaking anxiety [in Chinese]. *English Square*, (35), 112–115. https://doi.org/10.16723/j.cnki.yygc.2023.35.008

General Office of the Ministry of Education. (2007, July 10). *Notice on the issuance of the College English Curriculum Requirements (No. 3 [2007] of the General Office of the Ministry of Education)* [in Chinese]. http://www.moe.gov.cn/srcsite/A08/s7056/200707/t20070710_110825.html

Gu, Q., Fang, X., Chen, N., & Chen, F. (2025). The effects of strategy preparation time and anxiety level on state willingness to communicate [in Chinese]. *Foreign Language Studies*, *48*(5), 90–98+159.

Guo, J., & Liu, S. (2007). How to overcome the effects of classroom anxiety on English learning [in Chinese]. *Journal of China Youth University of Political Studies*, (2), 136–139. https://doi.org/10.16034/j.cnki.10-1318/c.2007.02.026

Guo, Y., & Xu, J. (2014). A multidimensional study of English learning anxiety among non-English-major college students [in Chinese]. *Foreign Language World*, (4), 2–11. https://doi.org/10.26971/j.cnki.flw.1004-5112.2014.04.001

Horwitz, E. K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, *21*, 112–126. https://doi.org/10.1017/S0267190501000071

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE.

Li, J., & Li, Q. (2016). Research on foreign language anxiety in China: A review and reflection based on a statistical analysis of papers in foreign language journals over the past decade (2006–2015) [in Chinese]. *Foreign Language World*, (4), 58–65. https://doi.org/10.26971/j.cnki.flw.1004-5112.2016.04.008

Li, P. (2014). Evaluation of a task-based teaching model in improving nursing students' anxiety about spoken nursing English and raising their achievement [in Chinese]. *Chinese Journal of Practical Nursing*, *30*(1), 7–10.

Liu, M. (2011). The effects of low self-confidence and classroom performance anxiety on college students' English learning: A cross-lagged study [in Chinese]. *Foreign Language Education*, *32*(5), 43–47. https://doi.org/10.16362/j.cnki.cn61-1023/h.2011.05.020

Luo, Y. (2019). A study of the relationship between foreign language anxiety and speaking participation among college students [in Chinese]. *Journal of Changchun Normal University*, *38*(11), 189–191.

Ma, L., Liu, L., & Huang, L. (2009). The effects of anxiety on oral output in foreign language learning: An exploration from the perspective of traditional culture [in Chinese]. *Modern Education Science*, (7), 111–113. https://doi.org/10.13980/j.cnki.xdjykx.gjyj.2009.07.005

MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, *82*(4), 545–562. https://doi.org/10.1111/j.1540-4781.1998.tb05543.x

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, *44*(2), 283–305. https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

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*, n71. https://doi.org/10.1136/bmj.n71

Qu, X. (2013). A review of research on English speaking anxiety in China [in Chinese]. *The Science Education Article Collects*, (29), 79+83.

Shi, Y., & Xu, J. (2013). Forty years of foreign language anxiety research at home and abroad: A statistical analysis of papers in 29 SSCI journals and 12 CSSCI journals over 40 years (1972–2011) [in Chinese]. *Foreign Languages and Their Teaching*, (1), 60–65. https://doi.org/10.13458/j.cnki.flatt.003868

Wang, J., & Wang, B. (2014). The effect of needs-analysis-based course design on reducing foreign language anxiety among non-English-major college students: An empirical study based on a college English public speaking course [in Chinese]. *Foreign Languages and Literature*, *30*(3), 157–162.

Wang, T. (2014). Creative applications of REBT techniques in treating foreign language speaking anxiety [in Chinese]. *Foreign Language Research*, (1), 125–128. https://doi.org/10.16263/j.cnki.23-1071/h.2014.01.027

Journal of Language
Volume 8, Number 2, pp: 373-389
e-ISSN: 2685-8878 | p-ISSN: 2655-9080
DOI: https://doi.org/10.30743/jol.v8i2.14496

Wang, Y. (2023). Exploring the relationship between L2 grit and willingness to communicate: The mediating effects of foreign language enjoyment and anxiety [in Chinese]. *Modern Foreign Languages, 46*(1), 42–55. https://doi.org/10.20071/j.cnki.xdwy.2023.01.012

Wu, R., & Zhao, C. (2013). The effects of pronunciation and intonation imitation training on the speaking anxiety of Chinese English learners [in Chinese]. *Journal of PLA University of Foreign Languages, 36*(5), 54–58.

Xiao, M. (2025). An empirical study of cooperative learning in alleviating higher vocational students' anxiety about learning spoken English [in Chinese]. *Journal of China Multimedia and Network Teaching (Mid-month)*, (1), 107–110.

Young, D. J. (1991). Creating a low-anxiety classroom environment: What does language anxiety research suggest? *The Modern Language Journal, 75*(4), 426–437. https://doi.org/10.1111/j.1540-4781.1991.tb05378.x

Yu, H., & Liu, T. (2021). Typical dynamic features of learners' speaking motivation: A retrospective paradigm approach [in Chinese]. *Foreign Languages and Their Teaching*, (4), 68–78+149. https://doi.org/10.13458/j.cnki.flatt.004792