ISELT 164
Interdisciplinary Studies in English Language Teaching

Pages 164 to 188

Comparing General and Scientometrics-Specific Academic Word Lists for Improving Reading Comprehension in Library and Information Science Students

Pooria Barzan [1], Reza Khany [1]*, Khalil Tazik [2]

[1] Ilam University, Ilam, Iran

[2] Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

10.22080/iselt.2026.29938.1117

Received
August 26, 2025
Accepted
May 7, 2026
Available online
May 13, 2026

Abstract

This study addresses a key gap in English for Specific Purposes (ESP) pedagogy within Library and Information Science (LIS) by developing and evaluating a Scientometrics-Specific Academic Word List (SSAWL) to enhance reading comprehension. General Academic Word Lists (AWLs), such as Coxhead's (2000), often overlook domain-specific terminology essential for scientometrics (e.g., "bibliometrics," "h-index"). A specialized 8.5-million-word corpus from leading scientometrics journals produced a validated SSAWL of 1,105 high-frequency word families. Using a quasi-experimental pretest–posttest design, 60 intermediate-to-advanced EFL undergraduate LIS students were assigned to an experimental group receiving 12 weeks of explicit SSAWL instruction via Research-based Academic Vocabulary Education (RAVE) or a control group receiving equivalent general AWL training. Assessments included multiple-choice, short-answer, and cloze reading tasks using authentic scientometrics passages, an adapted Vocabulary Levels Test, surveys, and focus groups. Results showed that the SSAWL group achieved significantly greater gains in scientometrics text comprehension (accuracy and speed) and in domain-specific vocabulary knowledge, with large effect sizes (Cohen's d = 1.67–3.85; partial η^2 = 0.39–0.79). Qualitative findings indicated improved confidence and contextual vocabulary use. This study demonstrates the superiority of corpus-derived, domain-specific word lists over general AWLs for specialized reading in LIS and offers a validated resource for ESP curriculum design.

Keywords:
Academic Word List, Scientometrics, Library and Information Science, Reading Comprehension, English for Specific Purposes (ESP).

1. INTRODUCTION

Reading comprehension is a cornerstone of academic success, particularly in specialized fields such as Library and Information Science (LIS), where students must navigate complex, domain-specific texts to engage with disciplinary knowledge effectively (Arboleda et al., 2022; Coxhead,

* **Corresponding Author:** Reza Khany, Professor, Department of English Language and Literature, Ilam University, Ilam, Iran, **Email**: r.khany@ilam.ac.ir

2000). Within LIS, the subfield of scientometrics, which focuses on quantitative study of science, scholarly communication, and research evaluation, presents unique lexical challenges as it relies on specialized terminology such as "bibliometrics," "h-index," and "citation analysis" (Hood & Wilson, 2001; Inastrilla et al., 2024; Sedighi, 2016). These terms, often underrepresented in general Academic Word Lists (AWLs) like Coxhead's (2000) seminal list, are critical for comprehending scientometric literature and participating in professional discourse. The inadequacy of general AWLs in addressing domain-specific lexical demands creates a significant barrier for English as a Foreign Language (EFL) LIS students, hindering their ability to process research articles efficiently and develop disciplinary expertise (Hyland & Tse, 2007; Tangsakul, 2024).

The critical role of vocabulary knowledge in reading comprehension is well-established in the literature, with both breadth (vocabulary size) and depth (understanding of word meanings and usage) serving as strong predictors of text understanding (Melani et al., 2021; Steensel et al., 2014; Tunmer & Chapman, 2012). General AWLs, designed to capture high-frequency academic vocabulary across disciplines, have been instrumental in English for Specific Purposes (ESP) pedagogy (Coxhead, 2000; Schmitt et al., 2011). However, empirical studies highlight that such lists often fail to account for the specialized lexis required in fields like scientometrics, leading to reduced lexical coverage and comprehension difficulties when students encounter domain-specific texts (Hyland & Tse, 2007; Lesaux et al., 2010; Vuković-Stamatović, 2024). This misalignment underscores the need for discipline-specific lexical resources, such as Scientometrics-Specific Academic Word Lists (SSAWLs), to bridge the gap between general academic vocabulary and the technical terminology essential for LIS education (Arboleda et al., 2022; Zhang, 2013).

Corpus-based methodologies have emerged as a robust approach to developing tailored word lists, enabling researchers to identify high-frequency, domain-relevant word families through systematic analysis of specialized texts (It-ngam & Phoocharoensil, 2019; Uehara et al., 2022). Prior studies in disciplines such as applied linguistics, physics, engineering, and medicine have demonstrated that specific academic word lists achieve superior lexical coverage and pedagogical relevance compared to general AWLs, supporting enhanced reading comprehension and disciplinary literacy (Chang, 2023; Khani & Tazik, 2013; Tazik et al., 2022; Yotimart, 2021). In scientometrics, where texts are characterized by dense technical jargon and methodological constructs, a corpus-derived Scientometrics-Specific Academic Word List (SSAWL) offers a promising solution to address the lexical demands of the field, aligning instructional materials with authentic reading tasks (Inastrilla et al., 2024; Luo et al., 2024).

Despite the recognized importance of vocabulary in reading comprehension, there is a paucity of empirically validated, domain-specific lexical resources for scientometrics within LIS education. General AWLs, while useful for broad academic contexts, often fail to capture the specialized terminology critical for understanding scientometric texts, potentially leading to inefficient vocabulary instruction and persistent comprehension challenges for EFL LIS students (Arboleda et al., 2022; Lesaux et al., 2010; Tangsakul, 2024). The lack of tailored lexical resources limits students' ability to engage with core scientometric literature, conduct literature reviews, and develop professional competencies, particularly in English-medium instruction (EMI) or ESP settings where lexical proficiency is paramount (Green et al., 2023; Rogobete et al., 2023). This gap necessitates the development and evaluation of a SSAWL to enhance reading comprehension and vocabulary knowledge in LIS programs.

The present study aims to develop and empirically evaluate the comparative efficacy of a SSAWL versus a general AWL in enhancing reading comprehension and domain-specific vocabulary knowledge among intermediate-to-advanced EFL undergraduate LIS students. Specific objectives include:

1. To compile a validated SSAWL, and compare the effectiveness of explicit SSAWL instruction, delivered through Research-based Academic Vocabulary Education (RAVE) methods, against equivalent general AWL instruction in a 12-week quasi-experimental intervention.
2. To assess improvements in reading comprehension and vocabulary knowledge using multi-method assessments, including authentic scientometrics passages, an adapted Vocabulary Levels Test, and qualitative data from surveys and focus groups.
3. To provide empirical evidence and a validated lexical resource to inform ESP curriculum design and instructional practices in LIS education.

This research addresses a critical gap in ESP pedagogy by providing a corpus-derived SSAWL tailored to the lexical demands of scientometrics, offering a replicable model for developing discipline-specific lexical resources. The findings contribute robust empirical evidence on the superiority of domain-specific vocabulary instruction, with implications for enhancing reading comprehension, vocabulary retention, and learner confidence in LIS programs. By aligning instructional materials with the authentic linguistic demands of scientometric texts, this research supports curriculum designers and educators in fostering disciplinary literacy and professional competence among EFL LIS students, particularly in EMI contexts (Chen et al., 2024; Xodabande et al., 2022). Additionally, the mixed-methods approach enriches the pedagogical discourse by integrating quantitative outcomes with qualitative insights into learners' experiences, informing future innovations in ESP vocabulary instruction (Hajiyeva, 2023).

In sum, this study advances the discourse on ESP vocabulary instruction by providing a validated SSAWL and empirical evidence for its efficacy in LIS education. By addressing the limitations of general AWLs and aligning vocabulary instruction with the lexical demands of scientometrics, the research offers a practical and theoretically grounded contribution to curriculum design, instructional practice, and disciplinary literacy development in specialized academic contexts.

2. LITERATURE REVIEW

Vocabulary Knowledge and Reading Comprehension

Vocabulary knowledge is widely recognized as a cornerstone of reading comprehension, serving as both a predictor and facilitator of academic literacy across proficiency levels and educational contexts. Empirical studies demonstrate that vocabulary breadth (the number of words known) and depth (semantic, morphological, and collocational knowledge) together explain a substantial proportion of variance in reading comprehension outcomes (Nation, 2006; Qian, 2002; Zhang & Yang, 2016). The lexical threshold hypothesis posits that readers must know approximately 95% of the words in a text for minimal comprehension and 98% for independent understanding (Laufer, 2013; Nation, 2006). In ESP, where texts are laden with technical and discipline-specific terms, lexical mastery becomes indispensable for comprehension and academic success (Chang, 2023; Masrai & Milton, 2018).

Limitations of General Academic Word Lists

Coxhead's (2000) Academic Word List, consisting of 570 word families, has been foundational for EAP pedagogy by identifying high-frequency academic vocabulary. However, subsequent research has revealed that general AWLs provide uneven coverage across disciplines and frequently omit critical domain-specific terms (Dang et al., 2017; Durrant, 2016; Hyland & Tse, 2007; Whitaker, 2024). For instance, scientometric concepts such as bibliometrics, h-index, and citation analysis, central to Library and Information Science, are absent or underrepresented in the AWL (Durrant, 2016; Gaffas, 2024; Sedighi, 2016). This gap underscores the need for discipline-specific lexical resources to better support learners engaging with specialized texts.

Development of Domain-Specific Word Lists

Corpus-based methodologies have emerged as the most reliable approach to developing domain-specific AWLs. By analyzing large, representative corpora, researchers can identify high-frequency items, collocations, and lexical bundles that define disciplinary discourse (Argyroulis, 2022; Vuković-Stamatović & Živković, 2022). Notable examples include the Applied Linguistics AWL (Khani & Tazik, 2013), Zoology AWL (Kruawong & Phoocharoensil, 2020), Forestry Word List (Luo et al., 2024), and Medical AWL (Tazik et al., 2022). These domain-specific lists consistently demonstrate superior lexical coverage and pedagogical relevance compared to Coxhead's AWL. For scientometrics, corpus-driven development of a SSAWL ensures the inclusion of technical vocabulary absent from general AWLs while reflecting the evolving nature of LIS discourse (Durrant, 2016; Sedighi, 2016; Whitaker, 2024).

Pedagogical Applications and Instructional Frameworks

Domain-specific vocabulary instruction has been shown to significantly enhance comprehension accuracy, speed, and retention compared to instruction based on general AWLs (Lawrence et al., 2021; McKeown et al., 2018). Effective frameworks, such as RAVE, integrate retrieval, automaticity, elaboration, and orthography, enabling learners to consolidate specialized vocabulary knowledge (Wolf et al., 2000, 2009). Studies highlight additional benefits, including improved learner confidence, engagement, and disciplinary literacy (Whitaker, 2024; Wright & Cervetti, 2016).

Technology-enhanced approaches further expand instructional possibilities. Digital tools such as concordancers, mobile applications, and game-based platforms provide authentic, contextualized exposure to specialized lexis, improving motivation and long-term retention (Xodabande et al., 2022). Integrating authentic scientometric texts, such as journal articles and bibliometric reports, into instruction strengthens the relevance and applicability of specialized vocabulary learning (Durrant, 2016; Sedighi, 2016).

Assessment and Methodological Considerations

Assessment of domain-specific AWL interventions requires multiple measures to capture both vocabulary growth and reading comprehension gains. Standardized instruments such as the Vocabulary Levels Test and comprehension tasks (e.g., cloze, multiple-choice, short-answer) are commonly used (Beglar, 2009; Schmitt et al., 2011). Pretest–posttest quasi-experimental designs consistently demonstrate moderate to very large effect sizes for domain-specific AWL-based interventions (Lawrence et al., 2021; McKeown et al., 2018). Complementary qualitative evidence also suggests affective benefits, such as increased confidence and reduced anxiety, when learners engage with domain-specific vocabulary instruction (Binhomran & Altalhab, 2023).

Challenges and Research Gap in LIS

Despite promising results in other disciplines, LIS and scientometrics remain underrepresented in the development and validation of domain-specific academic word lists (Luo et al., 2024; Nekrasova-Beker, 2020). Existing instruction often relies on general AWLs, which inadequately address the specialized lexical demands of scientometrics. Moreover, challenges such as resourceintensive corpus compilation, the need for teacher training, and the rapid evolution of scientometric terminology pose barriers to effective implementation (Jay & Sato, 2018).

Conclusion and Rationale for the Present Study

The literature converges on four key points: (1) vocabulary knowledge is critical for academic reading comprehension, (2) general AWLs, while useful, lack coverage of specialized terminology, (3) corpus-driven domain-specific AWLs provide superior lexical coverage and learning outcomes, and (4) explicit, technology-supported instruction maximizes their pedagogical impact. However, a significant gap remains in LIS, where no empirically validated SSAWL exists for scientometrics. The present study addresses this gap by developing a Scientometrics-Specific Academic Word List (SSAWL) and empirically evaluating its efficacy against a general AWL in enhancing LIS students' reading comprehension.

3. METHODOLOGY AND DESIGN

This study employed a mixed-methods approach to develop and evaluate the Scientometrics-Specific Academic Word List and its impact on reading comprehension among Library and Information Science students. The methodology was structured in two primary phases: the development of the SSAWL and the implementation of a quasi-experimental intervention. Data collection and analysis integrated quantitative and qualitative techniques to provide comprehensive insights into the efficacy of the SSAWL compared to a general Academic Word List.

Design

A quasi-experimental pretest-posttest design with a control group was utilized to assess the comparative effectiveness of SSAWL instruction versus general AWL training. This design facilitated causal inferences in a naturalistic educational setting, where full randomization was limited by existing class structures. The unit of assignment was intact classes (i.e., pre-existing student groups rather than individual students), with two classes randomly assigned to either the experimental or control condition after stratification to balance English proficiency (based on Oxford Placement Test scores) and prior LIS exposure (self-reported semesters completed). Stratification involved dividing students into strata based on these variables and ensuring proportional representation in each group. To minimize contamination risk, classes were scheduled in different time slots within the same term and department, with no shared activities between groups. Both groups were taught by the same instructor, who followed standardized lesson plans, materials, and pacing to ensure consistency across conditions. The experimental group received targeted SSAWL instruction, while the control group underwent equivalent training with a general AWL (Coxhead, 2000). The intervention spanned 12 weeks, allowing for pre- and post-intervention assessments to measure changes in reading comprehension and vocabulary knowledge.

Participants

A convenience sample of 60 undergraduate LIS students was recruited from Tehran University in an EFL context, with participants demonstrating intermediate to advanced English proficiency (equivalent to CEFR B2-C1 levels, as determined by a preliminary Oxford Placement Test). The sample size was determined through a priori power analysis using G*Power software, yielding sufficient power (>0.80) to detect medium effect sizes (Cohen's d = 0.50) at $\alpha = 0.05$, consistent with prior vocabulary intervention studies (Xodabande et al., 2022). Inclusion criteria required enrollment in LIS courses with scientometrics content, while exclusion criteria eliminated students with native-level English proficiency or prior specialized scientometrics training. Demographic balancing was achieved through stratification, resulting in comparable groups in terms of age (M = 21.4 years, SD = 1.8), gender (52% female), and prior LIS coursework exposure. Participants were evenly divided between the experimental (n = 30) and control (n = 30) groups.

Phase 1: SSAWL Development

The SSAWL was developed using a corpus-based methodology to ensure empirical validity and domain specificity. A specialized corpus comprising 8.5 million words was compiled from 1,800 full-text articles sourced from six highest impact factor scientometrics journals: *Scientometrics*, *Journal of Informetrics*, *Journal of the Association for Information Science and Technology (JASIST)*, *Journal of Information Science*, *Research Evaluation*, and *COLLNET Journal of Scientometrics and Information Management*. Articles were selected from publications spanning 2000–2024 to capture contemporary terminology (except for *COLLNET Journal of Scientometrics and Information* Management, which was first published in 2007), with texts processed to remove non-linguistic elements (e.g., references, tables) using AntConc (Anthony, 2024) software.

Tokenization and lemmatization were performed using AntWordProfiler's default settings, which reduce words to their base forms (e.g., "citations" to "citation"). Word families were defined following Bauer and Nation's (1993) Level 6 criteria, including the headword, inflections, and common derivations (e.g., "cite" family includes "citation," "cited," "citing"). Proper nouns (e.g., journal names) and abbreviations (e.g., "JASIST") were excluded unless they functioned as common lexical items (e.g., "h-index" retained as a hyphenated single token). Hyphenated terms were treated as unified tokens if domain-specific (e.g., "co-citation"). Word families were identified and selected solely based on their frequency, range, and dispersion in the scientometrics corpus using AntWordProfiler (Anthony, 2024b). Families from the GSL and AWL were included in the SSAWL if they met the domain-specific selection criteria (token frequency >28 per million words, range >50%, dispersion Juilland’s D >0.80). This inclusive approach defines the SSAWL as a comprehensive domain list (not an add-on excluding general words) that captures all high-frequency lexis relevant to scientometrics, whether general or specialized; overlaps with GSL/AWL are explicitly marked in Appendix B. Frequency analysis was conducted with AntWordProfiler (Anthony, 2024b) to extract word families, prioritizing metrics such as token frequency (>28 occurrences per million words), range (appearance in >50% of texts), and dispersion (Juilland's D >0.80 for even distribution). AWL and the GSL (West, 1953) families were not excluded from SSAWL if they met the word selection criteria. This yielded an initial list of 1,309 word families, which was refined to the top 1,105 through iterative filtering. Validation involved expert review by three LIS faculty members specializing in scientometrics, who rated items for relevance and pedagogical utility on a 5-point Likert scale (inter-rater reliability: κ = 0.85). Adjustments were made based on feedback, resulting in a finalized SSAWL.

Figure 1: Word cloud of 50 most frequent word families in SSAWL

Coverage analysis showed that the SSAWL provided an additional 11.2% lexical coverage in independent scientometric test texts (e.g., recent articles not in the corpus) beyond the GSL + AWL, compared to 4.5% for the general AWL alone. See Appendix for detailed software settings (e.g., AntConc stopword lists, AntWordProfiler range/dispersion parameters), full word family criteria, and the complete wordlist items.

Phase 2: Intervention

The 12-week intervention was delivered in-person during regular class sessions, with each session lasting 2 hours to maintain consistency across groups. Instruction for both groups was facilitated by the same trained ESP instructor to minimize instructor bias, and materials were standardized in duration and intensity.

The experimental group focused on explicit SSAWL instruction, introducing approximately 25 word families per week through RAVE strategies (Lesaux et al., 2010). Activities included:

- Morphological analysis (e.g., breaking down prefixes and suffixes in terms like "co-citation" or "altmetrics") and definitional instruction using bilingual glossaries.
- Contextual embedding via authentic scientometrics readings (e.g., journal abstracts and excerpts), supplemented by interactive exercises such as digital flashcards for spaced repetition (Anki software) and sentence-completion tasks.
- Collaborative activities, including group discussions and role-plays simulating scientometric analyses, to promote incidental learning and deeper retention through multiple exposures.

To address overlap with the general AWL (see Appendix B for shared items marked as italicized), instruction and materials for the experimental group highlighted domain-specific applications of overlapping terms within scientometrics contexts, while introducing the unique SSAWL terms not present in the general AWL. The control group received instruction exclusively on the general AWL items without scientometrics-specific contextualization. This design allowed evaluation of the added value of domain-specific focus and unique terms.

The control group received parallel instruction using Coxhead's (2000) general AWL, with equivalent activities adapted to non-domain-specific academic vocabulary (e.g., "analyze," "evaluate"). Fidelity was ensured through session observations and instructor logs, with attendance rates exceeding 95% in both groups.

Data Collection

A multi-method data collection strategy was implemented to triangulate findings, encompassing quantitative assessments and qualitative insights.

Pretest and posttest measures were administered to both groups at Weeks 1 and 13, respectively:

- Reading comprehension was evaluated using three authentic scientometrics passages (500–800 words each, sourced from the corpus journals), followed by 20 items per passage (10 multiple-choice, 5 short-answer, 5 cloze with SSAWL deletions to assess inference). Speed was operationally defined as words per minute (WPM), calculated as total words across passages divided by total time in minutes; participants read silently, with timing starting when they began the first passage and stopping upon completing the last item, using a digital stopwatch for precision. Short-answer items were scored using a 2-point rubric (0 = incorrect/irrelevant, 1 = partially correct with key elements, 2 = fully correct and complete), with two independent raters achieving inter-rater reliability (intraclass correlation coefficient, ICC = 0.92); discrepancies were resolved through consensus. Cloze items targeted SSAWL words in novel contexts not used in instruction, measuring vocabulary integration via contextual inference rather than direct recall; acceptable synonyms or semantically equivalent responses were scored correct to avoid teaching-to-the-test bias. Scoring emphasized accuracy (percentage correct) and speed (time to completion), with internal consistency (Cronbach's $\alpha = 0.87$).
- Vocabulary knowledge was assessed via an adapted Vocabulary Levels Test (Schmitt et al., 2001), targeting SSAWL items through definition-matching (receptive knowledge), sentence-completion (productive recall), and recognition tasks (30 items total; $\alpha = 0.82$).

  Supplementary data included:
- Self-reported surveys (Likert-scale, 15 items) administered post-intervention to gauge perceived comprehension efficacy and vocabulary confidence ($\alpha = 0.89$).
- Focus group interviews (semi-structured, 45 minutes each) with 15 participants (three focus group interviews, 5 participants each, randomly selected from the experimental group) explored qualitative insights into experiences, challenges, benefits, and perceived improvements in applying SSAWL terms to scientometric texts, audio-recorded and transcribed for analysis.

All instruments were piloted with a similar cohort ($n = 20$) to refine clarity and reliability, with counterbalancing applied to mitigate order effects.

Data Analysis

Quantitative data were analyzed using SPSS version 28.0. Descriptive statistics summarized participant characteristics and score distributions. Within-group changes were examined via paired-samples t-tests to detect pre-to-post improvements. Between-group differences were assessed with independent t-tests, ANOVA, or ANCOVA (covarying pretest scores and baseline proficiency). Effect sizes were calculated using Cohen's d (for t-tests) and partial η^2 (for ANOVA), with interpretations following conventions (e.g., $d \geq 0.50$ indicating moderate effects). Gain-score Cohen's d was computed to illustrate the magnitude of within-group changes, justified here as it highlights practical improvements in a pre-post design, while ANCOVA partial η^2 and adjusted mean differences provide between-group comparisons controlling for baselines (Plonsky & Oswald, 2014). The primary (headline) analysis for between-group differences was ANCOVA on posttest scores covarying pretest scores; this was selected because it directly adjusts for baseline differences in this quasi-experimental pretest-posttest design, provides clear partial η^2 effect sizes, and is the recommended approach for such studies to maximize interpretability and statistical power. Exploratory repeated-measures ANOVA (Time × Group) and regression models yielded congruent results and are not reported in detail here to maintain a clear primary analysis path and avoid any impression of method shopping.

Qualitative data from surveys (open-ended responses) and focus groups were analyzed thematically using NVivo software. An inductive coding approach identified emergent themes (e.g., "increased confidence in contextual application"), with inter-coder reliability ($\kappa = 0.82$) achieved through double-coding of 20% of transcripts. Triangulation integrated quantitative results with qualitative patterns to provide nuanced interpretations of the SSAWL's efficacy.

Ethical Considerations

The research strictly complied with ethical standards for studies involving human participants, prioritizing their well-being and the integrity of the data. Participation was voluntary, with informed consent obtained from all individuals before their involvement. Participants were fully informed about the study's objectives, procedures, potential benefits, and their right to withdraw at any time without consequences. To protect privacy, personal identifiers were replaced with unique codes to anonymize data, and confidentiality was upheld throughout the entire research process, from data collection to reporting.

4. RESULTS

Quantitative Findings

Descriptive statistics for pre- and post-intervention measures are presented in Table 1. Both groups showed improvements across all measures, but the experimental group (SSAWL instruction) exhibited substantially larger gains.

Table 1: Descriptive Statistics for Vocabulary, Reading Accuracy, and Speed by Group and Time (N = 60)

Measure	Time	Control Group (n=30)		Experimental Group (n=30)		Mean Difference (Post-Pre)	
		Mean	SD	Mean	SD	Control	Experimental
Vocabulary	Pre	14.44	2.65	15.04	2.93		
	Post	17.25	3.11	24.00	3.20	+2.81	+8.96
Accuracy (%)	Pre	59.05	10.07	60.66	10.35		
	Post	70.44	9.98	84.99	8.49	+11.39	+24.33
Speed	Pre	147.13	22.78	155.64	15.28		
	Post	168.20	25.43	192.54	18.63	+21.07	+36.90

Note: Vocabulary scored 0-30; Accuracy 0-100%; Speed in words per minute (WPM). Higher speed values indicate faster reading. Standard deviations are in parentheses.

Paired-samples t-tests revealed significant within-group improvements for both groups on all measures (all p < .001). For the control group: vocabulary (t(29) = 11.05), accuracy (t(29) = 13.17), and speed (t(29) = 12.56). For the experimental group: vocabulary (t(29) = 27.00), accuracy (t(29) = 17.72), and speed (t(29) = 20.04).

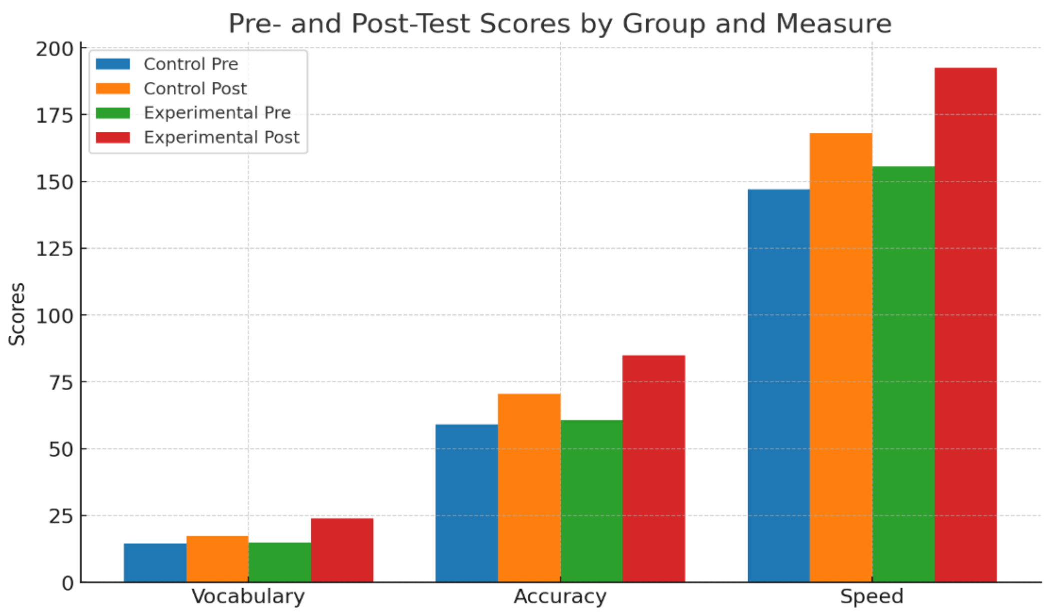

Figure 2: Pre- and Posttest Scores by Group and Measure

To assess between-group differences, ANCOVA was conducted on posttest scores, covarying pretest scores. Results are summarized in Table 2. Significant group effects were found for vocabulary (F (1,57) = 210.59, p < .001, η_p^2 = 0.79), accuracy (F (1,57) = 83.06, p < .001, η_p^2 = 0.59), and speed (F (1,57) = 36.47, p < .001, η_p^2 = 0.39), indicating superior outcomes for the

SSAWL group after controlling for baseline performance. Pretest scores were significant covariates in all models (all $p < .001$). Adjusted posttest means (controlling for pretest) were: vocabulary (experimental: 26.47, control: 17.53), accuracy (experimental: 89.12%, control: 64.88%), and speed (experimental: 184.03 WPM, control: 147.97 WPM).

Table 2: ANCOVA Results for Posttest Scores, Covarying Pretest Scores

Measure	Source	SS	df	F	p-value	$\eta p2$
Vocabulary	Group	561.92	1	210.59	<0.001	0.79
	Pretest	445.39	1	166.92	<0.001	
Accuracy	Group	2667.01	1	83.06	<0.001	0.59
	Pretest	3320.81	1	103.42	<0.001	
Speed	Group	3431.91	1	36.47	<0.001	0.39
	Pretest	24444.20	1	259.78	<0.001	

Cohen's *d* for gain scores further quantified the magnitude of differences: vocabulary (d = 3.85), accuracy (d = 2.09), and speed (d = 1.67), all representing large effect sizes.

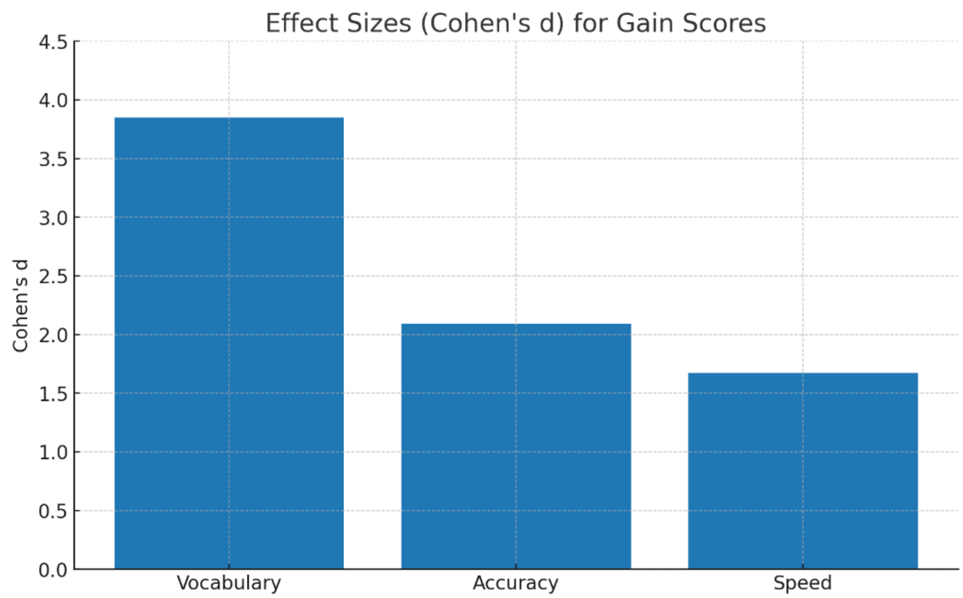

Figure 3: Effect sizes (Cohen's d) for gain scores

Qualitative Findings

Thematic analysis of focus group interviews (n = 15 from the experimental group) and open-ended survey responses identified three key themes illustrating the perceived benefits of SSAWL instruction:

1. Enhanced Domain-Specific Understanding: Participants reported that learning SSAWL terms like "h-index" and "citation analysis" facilitated deeper comprehension of scientometrics texts. One student stated, "*Before, words like 'bibliometrics' confused me, but now I understand the concepts behind them.*"

2. Increased Confidence in Reading Specialized Texts: Many noted greater self-assurance when engaging with authentic materials. A common sentiment was, "*I feel more confident reading journal articles because I know the key vocabulary.*"

3. Improved Application in Contextual Tasks: Students highlighted the ability to apply new vocabulary in discussions and assignments. For instance, "*Using terms like 'altmetrics' in my essays makes my analysis more professional.*"

These themes were prevalent across responses, with inter-coder reliability $\kappa = 0.82$.

Table 3: Summary of qualitative themes (SSAWL focus groups and open responses, n = 15)

Theme	Frequency (of 15 participants)	Description	Exemplary (anonymized) quote
Enhanced Domain-Specific Understanding	12/15	Learners described better ability to map SSAWL items to real research tasks (e.g., identifying h-index usage).	"When we saw 'co-citation', I could connect it to what we learned and explain why it mattered in the study."
Increased Confidence in Reading Specialized Texts	10/15	Participants reported feeling more able to approach abstracts, methods, and results sections, and to infer meaning from context.	"I could read an abstract and understand it without constantly checking the glossary; I felt more confident in class discussions."
Improved Application in Contextual Tasks	7/15	Several students reported using SSAWL items in assignments and in oral presentations, indicating productive uptake.	"I used 'altmetrics' and 'citation network' in my presentation; the instructor commented on the precision of my language."

Integrated Findings

The quantitative results demonstrate the superior efficacy of SSAWL over general AWL in enhancing vocabulary knowledge, reading accuracy, and speed, with large effect sizes. Qualitative insights complement these findings by elucidating how domain-specific vocabulary fostered confidence and practical application, aligning with the principles of ESP pedagogy. Together, the data provide strong evidence for the value of tailored lexical resources in LIS education.

5. DISCUSSION AND CONCLUSION

The findings of this study provide compelling evidence for the superior efficacy of a Scientometrics-Specific Academic Word List over a general Academic Word List in enhancing reading comprehension and domain-specific vocabulary knowledge among EFL undergraduate Library and Information Science students. The experimental group, which received explicit instruction in the SSAWL through RAVE methods, demonstrated significantly greater improvements in vocabulary knowledge (M gain = +8.96), reading accuracy (M gain = +24.33%), and reading speed (M gain = +36.90 words per minute) compared to the control group, with large effect sizes (Cohen's d ranging from 1.67 to 3.85, partial η^2 ranging from 0.39 to 0.79). These quantitative outcomes were corroborated by qualitative insights, where participants reported enhanced domain-specific understanding, increased confidence in engaging with specialized texts, and improved ability to apply SSAWL terms in contextual tasks. Such results affirm the study's primary hypothesis that domain-specific lexical resources better address the unique terminological demands of scientometrics, thereby facilitating more effective comprehension of disciplinary texts.

These results align closely with established literature on the limitations of general AWLs and the advantages of corpus-derived, discipline-specific word lists. As noted by Hyland and Tse (2007) and Dang et al. (2017), general AWLs like Coxhead's (2000) often fail to capture the specialized lexis essential for fields such as scientometrics, leading to inadequate coverage and comprehension barriers. The SSAWL's focus on high-frequency terms absent from general lists, such as "bibliometrics," "h-index," and "citation analysis", directly mitigated these gaps, resulting in superior lexical coverage and pedagogical relevance, consistent with findings from domain-specific lists in applied linguistics (Khani & Tazik, 2013), medicine (Tazik et al., 2022), and forestry (Luo et al., 2024). The large effect sizes observed here are consistent with those reported in similar interventions (e.g., Lawrence et al., 2021; McKeown et al., 2018), underscoring the value of tailoring vocabulary instruction to the dense technical jargon characteristic of scientometric discourse (Hood & Wilson, 2001; Inastrilla et al., 2024; Sedighi, 2016). Furthermore, the integration of RAVE strategies, which emphasize morphological analysis, contextual embedding, and spaced repetition, enhanced retention and transfer, echoing the benefits of explicit, multi-exposure approaches documented by Lesaux et al. (2010) and Whitaker (2024).

Qualitatively, the emergent themes of enhanced understanding, confidence, and application resonate with prior research highlighting the affective and practical dimensions of domain-specific vocabulary learning. For instance, participants' reports of reduced confusion with terms like "co-citation" and "altmetrics" mirror the improved disciplinary literacy and engagement noted in ESP studies (Wright & Cervetti, 2016). This increased self-assurance aligns with the lexical threshold hypothesis (Laufer, 2013; Nation, 2006), as greater familiarity with specialized lexis enabled learners to surpass the 95–98% coverage needed for independent comprehension, thereby reducing anxiety and fostering productive use in assignments and discussions (Binhomran & Altalhab, 2023). The triangulation of quantitative gains with these qualitative insights strengthens the argument that SSAWL instruction not only boosts measurable outcomes but also cultivates a more holistic disciplinary competence, particularly in English-medium instruction contexts (Chen et al., 2024; Xodabande et al., 2022).

The implications of these findings are multifaceted for ESP pedagogy and LIS curriculum design. Practically, the validated SSAWL offers a ready-to-use lexical resource that educators can integrate into courses, potentially through digital tools like Anki for personalized learning or concordancers for authentic exposure. This study advocates for a shift from reliance on general AWLs to customized interventions, providing a replicable model for other LIS subfields or disciplines with similar lexical demands (Argyroulis, 2022; Vuković-Stamatović & Živković, 2022). Theoretically, it contributes to the discourse on vocabulary's role in reading comprehension by demonstrating how domain-specific lists bridge the gap between breadth and depth of knowledge (Nation, 2006; Qian, 2002), thus informing broader frameworks for specialized literacy development (Chang, 2023; Masrai & Milton, 2018).

Despite these strengths, several limitations of the present study should be acknowledged. First, the sample was a convenience cohort of 60 undergraduate LIS students drawn from a single university in an EFL setting (Tehran), all at intermediate-to-advanced proficiency (CEFR B2–C1). Although stratified assignment and an a priori power analysis strengthened internal validity, the demographic and institutional homogeneity of the sample constrains the generalizability of the findings to learners from different educational contexts, L1 backgrounds, proficiency levels, or institutions. Relatedly, the intervention was delivered within regular class time by a single trained instructor. While using the same instructor reduced variability between conditions, it raises the possibility of instructor-related or expectancy effects; subtle differences in enthusiasm, rapport, or delivery could have contributed to the observed advantage for the SSAWL group despite efforts to standardize materials and observe fidelity. Second, the study examined outcomes immediately after a 12-week intervention and therefore cannot speak to the durability of the observed gains. Large posttest improvements in vocabulary, accuracy, and speed are promising, but without delayed post testing (e.g., at three or six months), the extent to which these gains persist, consolidate, or decay remains unknown. Future research incorporating longitudinal follow-ups would clarify retention and the need for continued reinforcement to maintain disciplinary lexical competence. Third, although the SSAWL was developed using a large, carefully curated corpus and validated by expert raters, the decisions involved in list construction, journal selection (high-impact scientometrics titles, 2000–2024), frequency thresholds, range and dispersion cutoffs, and the exclusion of general service items, reflect a set of methodological choices that shape the final list. Different corpus parameters or inclusion criteria might yield slightly different item sets; thus, the SSAWL presented here should be understood as one empirically defensible operationalization of scientometrics-relevant lexis rather than the sole or definitive list. Similarly, the comprehension assessment relied on three authentic passages drawn from the same or similar journals; while this choice enhances ecological validity, it also raises the possibility of some teaching-to-test alignment between instructional texts and assessment materials. Fourth, the qualitative component, focus groups and open responses, provided valuable insight into learners' experiences but was limited to participants from the experimental group. Including comparative qualitative data from the control group would have offered a fuller account of differential affective and metacognitive responses to the two instructional approaches. Finally, although class fidelity was high (attendance >95%) and instruments were piloted, the controlled, instructor-led nature of the intervention may not mirror implementation at scale or in asynchronous, resource-constrained settings. Practical constraints such as teacher training, availability of corpus-informed materials, and institutional curriculum requirements could influence outcomes in wider applications. These limitations do not invalidate the study's central findings but delimit their scope. The strengths of the design—mixed methods, rigorous corpus development, validated instruments, and robust effect sizes—support the conclusion that SSAWL instruction can substantially improve scientometrics reading performance in the studied context. Nonetheless, replication with more diverse samples, delayed retention testing, comparative qualitative sampling, and alternative corpus construction strategies will be important next steps to confirm robustness and guide scalable implementation.

In conclusion, this mixed-methods quasi-experiment provides strong empirical support for the pedagogical value of a corpus-derived SSAWL taught through RAVE-consistent methods. Compared with matched general AWL instruction, SSAWL learners achieved substantially larger gains in domain-specific vocabulary knowledge, reading accuracy, and reading speed, and reported heightened confidence and productive use of scientometrics terms. While limitations, especially related to generalizability and long-term retention, suggest caution, the study offers a clear proof of concept: aligning lexical targets to disciplinary text demands and teaching them explicitly using evidence-based procedures meaningfully enhances disciplinary reading comprehension. ESP practitioners, curriculum designers, and LIS educators should consider integrating domain-specific word lists and corresponding instructional designs into courses that prepare students to read, interpret, and produce scientometric research.

Declaration

During the preparation of this work, the authors used ChatGPT to increase the readability of the manuscript and improve its writing quality. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

Anthony, L. (2024a). *AntConc* (Version 4.3.1) [Computer software]. Waseda University. https://www.laurenceanthony.net/software/antconc/

Anthony, L. (2024b). *AntWordProfiler* (Version 2.2.1) [Computer software]. Waseda University. https://www.laurenceanthony.net/software/antwordprofiler/

Arboleda, G., Zambrano, M., Castillo, L., & Santos, J. (2022). Achieving reading comprehension through integrated vocabulary instruction. *Research Society and Development*, *11*(11), e197111133528. https://doi.org/10.33448/rsd-v11i11.33528

Argyroulis, V. (2022). Investigating student motivation in the use of corpus concordancing in ESP learning at university level. *Esp Today*, *10*(1), 71-98. https://doi.org/10.18485/esptoday.2022.10.1.4

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253–279

Beglar, D. (2009). A Rasch-based validation of the vocabulary size test. *Language Testing*, *27*(1), 101-118. https://doi.org/10.1177/0265532209340194

Binhomran, K. & Altalhab, S. (2023). A systematic review of EFL research on the impact of reading comprehension on students' vocabulary retention, and the relation between reading and vocabulary size. *International Journal of Language and Literary Studies*, *5*(3), 146-168. https://doi.org/10.36892/ijlls.v5i3.1395

Chang, L. (2023). A corpus-based mechanical engineering academic word list. *International Journal of TESOL Studies*, *5*(3), 126-142. https://doi.org/10.58304/ijts.20230310

Chen, L., Chang, K., Wu, C., & Chen, S. (2024). A machine-based corpus optimization method for extracting domain-oriented technical words: An example of covid-19 corpus data. *Journal of Intelligent & Fuzzy Systems*, *46*(4), 9945-9956. https://doi.org/10.3233/jifs-236635

Coxhead, A. (2000). A new academic word list. *Tesol Quarterly*, *34*(2), 213-238. https://doi.org/10.2307/3587951

Dang, T., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, *67*(4), 959-997. https://doi.org/10.1111/lang.12253

Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, *43*, 49–61. https://doi.org/10.1016/j.esp.2016.01.004

Gaffas, Z. (2024). Learning medical terminology in an ESP medical course. *Australian Review of Applied Linguistics*, *48*(3), 535 – 558. https://doi.org/10.1075/aral.24070.gaf

Green, C., Pantelich, M., Barrow, M., Weerasinghe, D., & Daniel, R. (2023). Receptive vocabulary size estimates for general and academic vocabulary at a multi-campus Australian university. *Australian Review of Applied Linguistics*, *47*(2), 153-173. https://doi.org/10.1075/aral.21099.gre

Hajiyeva, K. (2023). From needs analysis to the establishment of a university vocabulary list for English majors. *Arab World English Journal*, *14*(2), 77-89. https://doi.org/10.24093/awej/vol14no2.6

Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, *52*, 291–314. https://doi.org/10.1023/A:1017919924342

Hyland, K. & Tse, P. (2007). Is there an "academic vocabulary"?. *Tesol Quarterly*, *41*(2), 235-253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Inastrilla, C., Vitón-Castillo, A., Santana, M., Vera, D., & García, T. (2024). Information skills applied to scientometrics: A literature systematic review. *Health Leadership and Quality of Life*, *3*, 512. https://doi.org/10.56294/hl2024.512

It-ngam, T. & Phoocharoensil, S. (2019). The development of science academic word list. *Indonesian Journal of Applied Linguistics*, *8*(3), 657-667. https://doi.org/10.17509/ijal.v8i3.15269

Jay, V. & Sato, Y. (2018). Creating an institution-specific science and engineering academic word list for university students. *The Journal of Asiatefl*, *15*(1), 148-166. https://doi.org/10.18823/asiatefl.2018.15.1.10.148

Khani, R. & Tazik, K. (2013). Towards the development of an academic word list for applied linguistics research articles. *Relc Journal*, *44*(2), 209-232. https://doi.org/10.1177/0033688213488432

Kruawong, T. & Phoocharoensil, S. (2020). Developing an English zoology academic word list: A corpus-based study. *Thoughts*, (2), 63-78. https://doi.org/10.58837/chula.thts.2020.2.3

Laufer, B. (2013). Lexical thresholds for reading comprehension: What they are and how they can be used for teaching purposes. *TESOL Quarterly*, *47*(4), 867-872. https://doi.org/10.1002/tesq.140

Lawrence, J. F., Knoph, R., McIlraith, A., Kulesz, P. A., & Francis, D. J. (2021). Reading comprehension and academic vocabulary: Exploring relations of item features and reading proficiency. *Reading Research Quarterly*, *57*(2), 669–690. https://doi.org/10.1002/rrq.434

Lesaux, N., Crosson, A., Kieffer, M., & Pierce, M. (2010). Uneven profiles: Language minority learners' word reading, vocabulary, and reading comprehension skills. *Journal of Applied Developmental Psychology*, *31*(6), 475-483. https://doi.org/10.1016/j.appdev.2010.09.004

Luo, G., Song, J., & Wu, J. (2024). The lexical profile of forestry academic texts: What does it take to understand a specialized discipline?. *Plos One*, *19*(12), e0315975. https://doi.org/10.1371/journal.pone.0315975

Masrai, A. & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, *31*, 44-57. https://doi.org/10.1016/j.jeap.2017.12.006

McKeown, M. G., Crosson, A. C., Moore, D., & Beck, I. L. (2018). Word knowledge and comprehension effects of an academic vocabulary intervention for middle school students. *American Educational Research Journal*, *55*(3), 572–616. https://doi.org/10.3102/0002831217744181

Melani, B., Willian, S., Apgrianto, K., & Lail, H. (2021). Vocabulary coverage and reading comprehension of university EFL learners. In *Proceedings of the Thirteenth Conference on Applied Linguistics (CONAPLIN 2020)* (pp. 65–71). Atlantis Press. https://doi.org/10.2991/assehr.k.210427.010

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review/ La Revue Canadienne Des Langues Vivantes*, *63*(1), 59-82. https://doi.org/10.3138/cmlr.63.1.59

Nekrasova-Beker, T. (2020). Vocabulary demands for engineering students studying English in Russia: comparing ESP course materials across three engineering disciplines. *Global Business Languages*, *20*, 159-171. https://doi.org/10.4079/gbl.v20.9

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, *52*(3), 513-536. https://doi.org/10.1111/1467-9922.00193

Rogobete, R., Mureşan, V., & Chitez, M. (2023). Academic word lists in English and Romanian: A corpus-based contrastive analysis. *Quaestiones Romanicae*, *10*(1), 373-386. https://doi.org/10.35923/qr.10.01.27

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, *95*(1), 26-43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the vocabulary levels test. *Language Testing*, *18*(1), 55–88. https://doi.org/10.1177/026553220101800103

Sedighi, M. (2016). Application of word co-occurrence analysis method in mapping of the scientific fields: The case of informetrics. *Library Review*, *65*(1–2), 52–64. https://doi.org/10.1108/LR-07-2015-0075

Steensel, R., Oostdam, R., Gelderen, A., & Schooten, E. (2014). The role of word decoding, vocabulary knowledge and meta-cognitive knowledge in monolingual and bilingual low-achieving adolescents' reading comprehension. *Journal of Research in Reading*, *39*(3), 312-329. https://doi.org/10.1111/1467-9817.12042

Tangsakul, S. (2024). Relationship between vocabulary size and TOEIC reading achievement among undergraduate students. *International Journal of Sociologies and Anthropologies Science Reviews*, *4*(4), 305-312. https://doi.org/10.60027/ijsasr.2024.3896

Tazik, K., Maniati, M., Afshar, M., & Haghighi, S. (2022). Developing a new academic wordlist for medical purposes – a viable tool for educators. *GMS journal for medical education*, *39*(1), Doc9. https://doi.org/10.3205/zma001530

Tunmer, W. & Chapman, J. (2012). The simple view of reading redux. *Journal of Learning Disabilities*, *45*(5), 453-466. https://doi.org/10.1177/0022219411432685

Uehara, S., Haraki, H., & McLean, S. (2022). Developing a discipline-specific corpus and high-frequency word list for science and engineering students in graduate school. *Vocabulary Learning and Instruction*, *11*(2), 57-68. https://doi.org/10.7820/vli.v11.2.uehara

Vuković-Stamatović, M. (2024). Creating and validating a corpus-based English academic word list for physics. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, *38*(1), 80–106. https://doi.org/10.1075/resla.22041.vuk

Vuković-Stamatović, M., & Bratić, V. (2022). Selecting ESP reading materials: Vocabulary suitability of science magazines for English for science teaching and learning. *AAA: Arbeiten Aus Anglistik Und Amerikanistik*, *47*(1), 109–124. https://www.jstor.org/stable/27204981

West, M. (1953). *A general service list of English words: With semantic frequencies and a supplementary word-list for the writing of popular science and technology*. Longman; Green.

Whitaker, A. (2024). Teaching vocabulary in meaningful and memorable ways: Explicit instruction in the English classroom. *Georgia Journal of Literacy*, *46*(2), 50-64. https://doi.org/10.56887/galiteracy.166

Wolf, M., Barzillai, M., Gottwald, S., Miller, L., Spencer, K., Norton, E., … & Morris, R. (2009). The RAVE-O intervention: Connecting neuroscience to the classroom. *Mind Brain and Education*, *3*(2), 84-93. https://doi.org/10.1111/j.1751-228x.2009.01058.x

Wolf, M., Miller, L., & Donnelly, K. (2000). Retrieval, automaticity, vocabulary elaboration, orthography (rave-o). *Journal of Learning Disabilities*, *33*(4), 375-386. https://doi.org/10.1177/002221940003300408

Wright, T. & Cervetti, G. (2016). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly*, *52*(2), 203-226. https://doi.org/10.1002/rrq.163

Xodabande, I., Iravi, Y., Mansouri, B., & Matinparsa, H. (2022). Teaching academic words with digital flashcards: Investigating the effectiveness of mobile-assisted vocabulary learning for university students. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.893821

Yotimart, D. (2021). Academic vocabulary in sport tourism news: A corpus-based study. *Journal of Language and Linguistic Studies*, *17*(3), 1527-1535. https://doi.org/10.52462/jlls.110

Zhang, D. & Yang, X. (2016). Chinese l2 learners' depth of vocabulary knowledge and its role in reading comprehension. *Foreign Language Annals*, *49*(4), 699-715. https://doi.org/10.1111/flan.12225

Zhang, M. (2013). The application of corpus tools in the teaching of discipline-specific academic vocabulary. *International Journal of Computer-Assisted Language Learning and Teaching*, *3*(4), 33-47. https://doi.org/10.4018/ijcallt.2013100104

Appendices

Appendix A: SSAWL Development Details

- *Software Settings:* AntConc (v4.3.1) used for corpus cleaning with default tokenization (whitespace-separated); AntWordProfiler (v2.2.1) with frequency threshold >28/million, range >50%, dispersion D>0.80, lemma list enabled.
- *Word Family Criteria:* Based on Bauer & Nation Level 6 (headword + inflections/derivations, e.g., cite: citation, cited, citing, citable).

Appendix B: SSAWL (Word families ranked by frequency)

Words shared with the AWL are *italicized*, those in the first 1,000 GSL are in regular font, those in the second 1,000 GSL are **bold**, and non-GSL/non-AWL words are underlined.

1.	*research*	*11.*	*publication*	21.	set	31.	quality
2.	*citation*	12.	science	22.	year	*32.*	*academic*
3.	*journal*	13.	study	**23.**	**model**	*33.*	*evaluation*
4.	number	*14.*	*analysis*	24.	high	34.	different
5.	*data*	15.	result	25.	social	35.	example
6.	paper	*16.*	*index*	26.	average	*36.*	*network*
7.	article	17.	web	*27.*	*indicator*	37.	collaboration
8.	*impact*	18.	field	28.	found	38.	use
9.	**information**	19.	knowledge	29.	country	39.	value
10.	*author*	20.	level	30.	total	40.	university

41. term
42. document
43. important
44. distribution
45. case
46. performance
47. method
48. related
49. review
50. figure
51. reference
52. factor
53. process
54. subject
55. system
56. literature
57. group
58. approach
59. international
60. similar
61. significant
62. topic
63. period
64. discipline
65. order
66. bibliometric
67. institution
68. individual
69. since
70. section
71. database
72. funding
73. according
74. measure
75. specific
76. self
77. consider
78. ranking
79. output
80. count
81. access
82. correlation
83. text
84. technology
85. compare
86. type
87. category
88. general
89. possible
90. main
91. influence
92. several
93. area
94. scholar
95. effect
96. mean
97. peer
98. sample
99. rate
100. development
101. list
102. classification
103. link
104. low
105. single
106. increase
107. available
108. particular
109. various
110. present
111. open
112. provide
113. applied
114. library
115. include
116. receive
117. growth
118. productivity
119. size
120. previous
121. selected
122. degree
123. content
124. define
125. source
126. world
127. variable
128. share
129. finding
130. metric
131. structure
132. identify
133. score
134. manage
135. overall
136. respectively
137. obtain
138. communication
139. policy
140. part
141. percentage
142. keyword
143. full
144. problem
145. pattern
146. assessment
147. cluster
148. need
149. page
150. fact
151. project
152. algorithm
153. context
154. common
155. nodes
156. future
157. domain
158. positive
159. calculate
160. recent
161. test
162. dataset
163. role
164. core
165. function
166. observe
167. current
168. focus
169. feature
170. public
171. expect
172. language
173. query
174. major
175. characteristic
176. computer
177. potential
178. education
179. account
180. change
181. online
182. activity
183. question
184. support
185. collection
186. engineering
187. multiple
188. criteria
189. learn
190. point
191. still
192. form
193. corresponding
194. associated
195. organizations
196. range
197. conference
198. evidence
199. production
200. records
201. instance
202. especial
203. report
204. patents
205. frequency
206. statistical
207. students
208. nature
209. physics
210. items
211. extent
212. sets
213. mention
214. standard
215. members
216. concept
217. certain
218. conduct
219. innovation
220. usage
221. contribution
222. title
223. hence
224. humanities
225. interest
226. coverage
227. long
228. law

229. aspect
230. issue
231. age
232. key
233. probability
234. strong
235. negative
236. describe
237. bibliographic
238. known
239. global
240. traditional
241. empirical
242. book
243. basic
244. must
245. framework
246. discuss
247. shape
248. view
249. proportion
250. scale
251. basis
252. retrieve
253. position
254. normalize
255. ratio
256. theory
257. limited
258. represent
259. decision
260. name
261. software
262. take
263. assign
264. participant
265. human
266. survey
267. original
268. clear
269. techniques
270. life
271. now
272. electronic
273. understanding
274. seems
275. become
276. reason
277. equal
278. task
279. note
280. people
281. next
282. weight
283. usually
284. practice
285. help
286. early
287. regression
288. attention
289. independent
290. medicine
291. local
292. frequently
293. call
294. matrix
295. relevance
296. furthermore
297. difficult
298. institute
299. class
300. determine
301. health
302. examine
303. tend
304. faculty
305. power
306. produce
307. together
308. step
309. existing
310. digital
311. centrality
312. involve
313. graph
314. quantitative
315. around
316. short
317. regarding
318. almost
319. appear
320. respondents
321. direct
322. accuracy
323. control
324. address
325. business
326. able
327. scientometrics
328. necessary
329. instead
330. suggest
331. trend
332. industry
333. tools
334. internet
335. purpose
336. half
337. behavior
338. simple
339. random
340. team
341. lack
342. window
343. bibliometrics
344. abstract
345. advantage
346. state
347. maximum
348. perspective
349. theoretical
350. bias
351. appropriate
352. classified
353. parameter
354. complex
355. external
356. line
357. unique
358. contain
359. mathematics
360. end
361. establish
362. gender
363. reflect
364. already
365. quite
366. yet
367. contrast
368. extract
369. media
370. technical
371. biology
372. complete
373. career
374. far
375. later
376. design
377. real
378. always
379. expert
380. length
381. cross
382. respect
383. map
384. effective
385. economic
386. space
387. amount
388. towards
389. actual
390. initial
391. improve
392. equation
393. procedure
394. wide
395. site
396. dimension
397. lead
398. environment
399. strategy
400. majority
401. element
402. diversity

403. pair
404. training
405. distance
406. multi
407. hypothesis
408. units
409. path
410. place
411. generate
412. introduce
413. target
414. alternative
415. proceeding
416. personal
417. editor
418. history
419. require
420. write
421. semantic
422. investigate
423. metadata
424. outcome
425. program
426. little
427. create
428. latter
429. stage
430. input
431. allow
432. evolution
433. broad
434. variety
435. service
436. significance
437. despite
438. government
439. combine
440. applicant
441. median
442. objective
443. indeed
444. reader
445. affect
446. staff
447. aim
448. grant
449. true
450. divide
451. choose
452. emerging
453. vector
454. society
455. intellectual
456. detail
457. idea
458. multidisciplinary
459. status
460. visibility
461. sense
462. along
463. carry
464. companies
465. united
466. material
467. prior
468. experience
469. sector
470. derive
471. component
472. department
473. success
474. version
475. interaction
476. comprehensive
477. close
478. dependent
479. entity
480. scheme
481. date
482. linear
483. experiment
484. physical
485. choice
486. popular
487. course
488. qualitative
489. population
490. scope
491. matching
492. property
493. consistent
494. altmetrics
495. fit
496. let
497. precision
498. strength
499. formal
500. proposal
501. entire
502. easily
503. except
504. oriented
505. cognitive
506. log
507. cooperation
508. professional
509. prediction
510. internal
511. percentile
512. explain
513. series
514. effort
515. explore
516. reputation
517. special
518. annual
519. slightly
520. threshold
521. situation
522. validity
523. employ
524. perceive
525. distinct
526. reveal
527. underlying
528. practical
529. cumulative
530. prestige
531. recall
532. professor
533. beyond
534. kind
535. remaining
536. rules
537. perhaps
538. phenomenon
539. presence
540. affiliation
541. foreign
542. primary
543. column
544. sentences
545. regional
546. differ
547. critical
548. engines
549. submit
550. credit
551. conclusion
552. achieve
553. conditions
554. domestic
555. typically
556. free
557. assumption
558. mobility
559. repository
560. corpus
561. accept
562. approximately
563. few
564. news
565. connect

566. concern
567. play
568. post
569. substantial
570. agreement
571. right
572. actors
573. efficiency
574. error
575. essential
576. consequent
577. fraction
578. locate
579. volume
580. downloads
581. goal
582. benefits
583. minimum
584. answer
585. occurrence
586. appendix
587. assume
588. decrease
589. estimate
590. diffusion
591. profile
592. panel
593. detection
594. machine
595. adopted
596. reliability
597. challenges
598. side
599. interpretation
600. enough
601. subsequent
602. response
603. colleagues
604. mining
605. summary
606. solution
607. market
608. exclude
609. agency
610. geographical
611. consist
612. female
613. cost
614. hybrid
615. transfer
616. formula
617. constant
618. phase
619. psychology
620. arts
621. belong
622. turn
623. generation
624. curve
625. inclusion
626. teaching
627. ask
628. dynamic
629. come
630. private
631. format
632. risk
633. mechanism
634. excellence
635. temporal
636. versus
637. overlap
638. considerable
639. south
640. price
641. million
642. implication
643. interview
644. integration
645. promotion
646. former
647. description
648. expressed
649. firms
650. look
651. financial
652. demonstrate
653. construct
654. rest
655. interval
656. say
657. left
658. believe
659. print
660. candidate
661. gold
662. objects
663. background
664. confidence
665. gain
666. doctoral
667. matter
668. fundamental
669. separate
670. skewed
671. stable
672. chemical
673. comment
674. nearly
675. suitable
676. comparative
677. political
678. sentiment
679. manually
680. cell
681. conclude
682. density
683. file
684. visualization
685. extend
686. clinical
687. interface
688. allocation
689. overview
690. argue
691. reduce
692. recommendation
693. expertise
694. missing
695. biomedical
696. letter
697. put
698. somewhat
699. obvious
700. implement
701. imply
702. gap
703. joint
704. month
705. event
706. robust
707. skills
708. attributes
709. build
710. seek
711. attempt
712. capture
713. insights
714. manuscript
715. analytical
716. sufficient
717. avoid
718. beginning
719. automatic
720. academia
721. trust
722. medium
723. ensure
724. flow
725. tags
726. partners
727. depth
728. recognize
729. advanced

730. disseminate
731. energy
732. draw
733. fixed
734. offer
735. discovery
736. questionnaire
737. novel
738. motivation
739. besides
740. thesis
741. absolute
742. delay
743. young
744. board
745. weak
746. commercial
747. reject
748. correct
749. exchange
750. decade
751. explicit
752. picture
753. theme
754. back
755. opportunity
756. serve
757. body
758. old
759. culture
760. aggregate
761. facilitate
762. home
763. informal
764. illustrate
765. restricted
766. indirect
767. couple
768. foundation
769. meta
770. middle
771. opinion
772. family
773. hard
774. aware
775. discourse
776. extensive
777. move
778. participation
779. rise
780. male
781. distinguish
782. plagiarism
783. download
784. sequence
785. capital
786. characterize
787. baseline
788. hierarchical
789. numerous
790. alone
791. principal
792. attract
793. council
794. light
795. denotes
796. directed
797. reach
798. ties
799. cause
800. molecule
801. evident
802. worth
803. incentives
804. competitive
805. continuous
806. exponential
807. none
808. agriculture
809. channel
810. remove
811. scarce
812. informetrics
813. confirm
814. capacity
815. direction
816. quartile
817. readership
818. progress
819. sociology
820. biochemistry
821. assistance
822. principle
823. prominent
824. deviation
825. edges
826. shift
827. crucial
828. decline
829. behind
830. check
831. deal
832. committee
833. phrases
834. think
835. enhance
836. subjective
837. emphasis
838. exception
839. manner
840. biological
841. edition
842. equivalent
843. proxy
844. extreme
845. concentrate
846. mode
847. retract
848. identical
849. instrument
850. school
851. consistency
852. consequence
853. false
854. filter
855. square
856. genetic
857. intend
858. face
859. utility
860. run
861. today
862. expansion
863. senior
864. exercise
865. code
866. legal
867. innovative
868. classic
869. modern
870. image
871. plan
872. balance
873. rapid
874. plot
875. feedback
876. translation
877. latent
878. credibility
879. regular
880. subscription
881. axis
882. label
883. automated
884. macro
885. prolific
886. century
887. enable
888. block
889. conventional
890. span
891. validation
892. return
893. notion

894. referee
895. upper
896. utilize
897. union
898. submission
899. cancer
900. college
901. driven
902. consensus
903. blog
904. operation
905. venue
906. ontology
907. optimal
908. session
909. demand
910. email
911. heterogeneous
912. chance
913. dissertation
914. inventor
915. keep
916. spread
917. bottom
918. platform
919. duplicate
920. fast
921. plus
922. pure
923. exact
924. solve
925. surprising
926. neither
927. earth
928. sensitive
929. tail
930. things
931. responsible
932. intention
933. proximity
934. wise
935. intelligence
936. peak
937. root
938. adjusted
939. city
940. origin
941. specify
942. net
943. stem
944. guideline
945. prefer
946. apparent
947. constitute
948. contrary
949. micro
950. treatment
951. initiatives
952. claim
953. proper
954. option
955. neural
956. abroad
957. artificial
958. intensity
959. lot
960. job
961. mixed
962. authority
963. row
964. award
965. double
966. boundary
967. dominant
968. binary
969. entries
970. meet
971. capabilities
972. engagement
973. north
974. press
975. infrastructure
976. absence
977. display
978. graduate
979. judgment
980. debate
981. white
982. competition
983. embedded
984. biotechnology
985. minor
986. elite
987. roughly
988. modified
989. opposite
990. barriers
991. copyright
992. day
993. try
994. forward
995. chapter
996. intrinsic
997. dependence
998. practitioners
999. really
1000. fair
1001. hyperlink
1002. parallel
1003. astronomy
1004. priority
1005. style
1006. webometrics
1007. complementary
1008. favor
1009. highlight
1010. decide
1011. investment
1012. loss
1013. rarely
1014. hot
1015. narrow
1016. postdoc
1017. yield
1018. track
1019. overcome
1020. pressure
1021. entropy
1022. simultaneous
1023. income
1024. prove
1025. speed
1026. transform
1027. undertaken
1028. fall
1029. mind
1030. moderate
1031. virtual
1032. laboratory
1033. quickly
1034. quantify
1035. marked
1036. headings
1037. care
1038. ease
1039. repeat
1040. outliers
1041. reward
1042. cloud
1043. host
1044. compose
1045. front
1046. literacy
1047. applicable
1048. coherence
1049. extension
1050. geometric
1051. color
1052. firm
1053. multidimensional
1054. predatory
1055. supervise
1056. mutual
1057. verify

1058. vocabulary
1059. audience
1060. maintain
1061. nanotechnology
1062. propensity
1063. rich
1064. adequate
1065. generic
1066. huge
1067. numerical
1068. promising
1069. attitude
1070. gathered
1071. fuzzy
1072. norms
1073. logic
1074. apart
1075. tenure
1076. transition
1077. corporate
1078. implicit
1079. string
1080. simulation
1081. logistic
1082. magnitude
1083. preliminary
1084. disease
1085. paid
1086. exclusive
1087. exhibit
1088. formation
1089. cycle
1090. monograph
1091. rights
1092. taxonomy
1093. division
1094. emphasize
1095. inherent
1096. null
1097. master
1098. bring
1099. registered
1100. ever
1101. seminal
1102. paradigm
1103. webometric
1104. mission
1105. sophisticated