

# Digital tools for a broad data-driven learning approach in mixed linguistic-proficiency ESP courses

Alejandro Blas Curado Fuentes 

acurado@unex.es

Universidad de Extremadura, Spain

Curado Fuentes, Alejandro Blas (2025). Digital tools for a broad data-driven learning approach in mixed linguistic-proficiency ESP courses. *Language Value*, 18(1), 18-48. Universitat Jaume I ePress: Castelló, Spain. <http://www.languagevalue.uji.es>.

July 2025

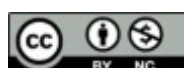
DOI: <https://www.doi.org/10.6035/languagev.8793>

ISSN 1989-7103

## ABSTRACT

Addressing the limited exploration of Data-Driven Learning (DDL) with mixed linguistic proficiencies in English for Specific Purposes (ESP), the purpose of this study is to experiment with a broad DDL (BDDL) approach in two English for business and tourism courses. The information collected through classroom activities, pre-, mid-, and post-tests, alongside polls and interviews, largely points to positive outcomes with BDDL at all linguistic levels. Lower-linguistic proficiency participants showed a greater challenge with linguistic analysis, but, in contrast with other corpus linguistics-based approaches in these ESP settings, BDDL worked more effectively to accommodate task procedures for these learners.

**Keywords:** *BDDL; DDL; ESP; Linguistic proficiency; Academic texts.*



## **I. INTRODUCTION**

Data-Driven Learning (DDL) emphasizes the personalization of linguistic analysis by actively involving learners in the discovery of linguistic patterns over digital and non-digital texts (Johns, 1991). For over three decades, DDL has been observed as especially beneficial in university contexts, promoting and heightening learners' autonomous work, problem-solving abilities, analytical skills, and language awareness (Boulton & Cobb, 2017).

However, the full integration of DDL requires its adoption by a broader educational community. Two variables noticeably under-researched in DDL are low linguistic proficiencies and languages for specific purposes (Boulton & Cobb, 2017; Boulton & Vyatkina, 2021; Boulton & Vyatkina, 2024; Boulton & Pérez-Paredes, 2025). In fact, university students outside language-related degrees and with less linguistic competence may feel more challenged and / or less involved with corpus analyses (Boulton, 2016). As a result, one possible approach to these varied contexts is the application of a BDDL (Broad DDL) (Pérez-Paredes, 2024) or "DDL-lite" perspective that may "bring DDL to the learners rather than requiring them to leap into corpus linguistics" (Boulton & Vyatkina, 2024, p. 1201). This BDDL perspective proposes not only simpler, user-friendly online tools but also linguistic analysis tailored to non-linguistics and / or non-humanities contexts. In this scope, linguistic explorations can be adapted according to English for Specific Purposes (ESP) students' particular needs. Additionally, because DDL seems to differ from other pedagogic approaches in terms of its reduced focus on communicative interactions between learners, among other aspects (Boulton, 2024), BDDL may offer a suitable option for emphasizing interactional / collaborative dynamics.

This article explores BDDL in two mixed linguistic-proficiency ESP courses, focusing on first-year students in the double degree of Business Management and Tourism at University of Extremadura, Spain. In these courses, BDDL has already been deployed over the past years in the form of simple-query online concordancers, text analytical

tools, online dictionaries, translators, and online videos. This study focuses on the first year of these studies because, at this stage, students are often required to engage with academic English texts in other subjects, and yet, most students lack the academic competence to comprehend the reading content effectively. Therefore, BDDL in these subjects has been specifically aimed at supporting mixed linguistic-level learners in working with academic texts, focusing on vocabulary and reading comprehension.

Learning progress, difficulties, and differences were analysed in these ESP courses by relying on mixed-method strategies, such as classroom observation, pre-, mid-, and post-intervention tests, post-task polls, and interviews. A main factor observed to influence learning developments was learners' linguistic proficiencies. One main occurrence was that lower-linguistic proficiency learners had more difficulties with linguistic analytical developments. However, compared to other DDL studies in similar ESP contexts (e.g., Curado Fuentes, 2015), BDDL showed a higher degree of success for empowering lower-linguistic level learners.

The following section presents a literature review that highlights the key theoretical and methodological foundations underpinning this study, followed by three research questions on which this study is based. The Methodology section then details the participants, course design, and evaluations conducted. The Results section describes the findings from both courses, while the Discussion highlights learning features, implications, and differences. Finally, the Conclusion offers a reflection on the main insights emerging from this case study.

## **II. LITERATURE REVIEW**

This section examines the applicability of BDDL to language learning. Also, because our case study focuses on lexical knowledge and text comprehension, some research directions are specified about vocabulary and reading in relation to DDL.

BDDL or DDL-lite (Boulton & Vyatkina 2024; Pérez-Paredes, 2024) includes, among other aspects, the use of familiar / user-friendly tools on the internet as a form of DDL-

like exploration for language learning. This scope connects with the use of NLP (Natural Language Processing) tools on the web, such as simple concordancers, translators, dictionaries, collocation finders, and so on (Ordoñana-Guillamón et al., 2024; Pérez-Paredes et al., 2018). Potential benefits of extending this broad conception of DDL across educational backgrounds include the promotion of linguistic awareness through autonomous interaction with various types of NLP resources (Ordoñana-Guillamón et al., 2024, p. 87). However, few studies have followed on this exploration of BDDL for diverse language learning goals and scenarios.

In BDDL, students exploit authentic linguistic input with a focus on the “L” of DDL (“Learning”) (Pérez-Paredes, 2024, p. 218), and not on the “D” (“Data”, e.g., corpus analyses). These tools can be aimed at enhancing specific linguistic learning, which constitutes an important aspect of digital developments in ESP (Bárcena et al., 2015; Xie, 2024). It is also relevant that these tools are directed at cultivating critical academic skills such as analytical and independent learning. These affordances can contribute to learners’ empowerment by allowing them to work at their own pace and to take control of their own learning experience, providing them with linguistic affordances that cater to their academic needs (Criollo et al., 2024). In this scope, creating dynamic digital environments that promote specific linguistic developments is important (Arce & Valdivia, 2020).

BDDL can incorporate the four stages of DDL (Carter & McCarthy, 1995; Flowerdew, 2009): 1) Illustration, i.e., explaining with linguistic examples; 2) Interaction, i.e., engaging with linguistic data and discussion; 3) Intervention, i.e., clarifying procedures and specifying directions; and 4) Induction, i.e., inferring patterns of linguistic use. An important aspect of BDDL is that these tools be “simple” and easily accessible (Ronan, 2023, p. 34), since more sophisticated corpus resources tend to intimidate less linguistically proficient students (Boulton, 2012; Timmis, 2015). Tools like concordance software packages and large corpora can also work well in ESP and English for Academic Purposes (EAP) settings, but more so with linguistically advanced students (Boulton, 2016; Charles, 2014).

BDDL can also encompass Generative Artificial Intelligence (GenAI) tools like ChatGPT, which students are increasingly using for linguistic queries (Boulton & Pérez-Paredes, 2025). English-as-a-foreign-language studies on learners' attitudes towards using ChatGPT reveal a spectrum of responses, ranging from enthusiasm to more reserved acceptance (Luo & Zuo, 2024). In ESP and EAP, these tools hold significant potential for fostering critical thinking, collaboration, and other academic skills (Karakose & Tülübas, 2023). They present both challenges and opportunities for the use of BDDL, since GenAI can offer instant linguistic input for contrastive analysis (Crosthwaite & Baisa, 2023). Thus, a crucial pedagogical approach can be the guiding of learners in using these tools to develop critical academic skills for language acquisition (Liu et al., 2024), for autonomous academic tasks like digital reading (Pan et al., 2024), and for vocabulary enhancement (Law, 2024), among other possibilities.

Lexical competence is actually an important aspect that BDDL can address, as vocabulary often constitutes a major difficulty in ESP due to specific word meanings (e.g., business vocabulary: Xie, 2024), whereas intermediate-level English learners' vocabulary is found to significantly increase with DDL tasks (Lee et al., 2019). DDL-related tactics can be adopted by focusing on pivotal academic lexical items (Argyroulis, 2022; Chambers, 2022; Hadley & Hadley, 2021; Lee et al., 2019). A BDDL approach can support this focus by relying on simple queries for linguistic pattern investigation, which allows students to detect key meanings within specific contexts (Curado Fuentes, 2024a; Pérez-Paredes et al., 2018; Phillip, 2010; Xu et al., 2019). Strategies such as noticing vocabulary by identifying "immediate collocates (and translation of the whole), [as well as] the value of example sentences, recognizing a word's syntactic role, using context to disambiguate, and so on" (Phillip, 2010, p. 12) are valuable. While broad vocabulary commands are beneficial, depth of lexical knowledge in specialized domains is crucial for reading comprehension (Song & Reynolds, 2022). This depth includes familiarity with word families, which often span multiple word classes, such as the verb *find* and the noun *findings* in specific contexts (Laufer & Cobb, 2019).

Because lexical complexity and specialization are often observed as a major factor of learners' difficulties with academic texts (Laufer & Ravenhorst-Kalovski, 2010), one noteworthy approach to cope with these problems is the encouragement of group-based strategies for the enhancement of familiarity with academic discourse (Flowerdew & Peacock, 2001; Basturkmen, 2010). Motivation is fostered when subject content-related texts are exploited by groups, since a sense of usefulness is perceived by ESP students when notions and ideas are discussed and shared (Nishizawa et al., 2018). Furthermore, meta-analyses of DDL developments reveal that peer collaboration works positively towards the integration of lower linguistic-proficiency learners using these tools (Lusta et al., 2023). In ESP, authentic linguistic tasks and group-based approaches should be designed around "problem-solving" (Belcher, 2009, p. 9) to meet academic learning needs (Nel, 2008), whereas collaboration can become a key strategy for problem-solving using digital tools (Tate & Warschauer, 2022).

Another issue is that, compared with academic writing, reading comprehension has been seldom addressed in DDL (Boulton & Cobb, 2017; Curado Fuentes, 2023; Hadley, 2025), and yet, DDL can be effectively explored for enhancing academic reading (Hadley & Hadley, 2021; Hadley, 2025; Lee et al., 2019). Text comprehension can be approached by focusing on specific corpus-driven vocabulary (Laufer & Ravenhorst-Kalovski, 2010; Paribakht & Webb, 2016), since, in DDL, decoding skills are enabled by noticing linguistic patterns within "repeated strings of text" (Cobb, 2018, p. 201). Additionally, the use of online tools in DDL can favour the design of effective mechanics for students' selection, compilation, and management of readings within their fields, fostering a sense of ownership over the texts (Charles, 2012). This ad-hoc corpus practice can enable learners to better understand the connection between specific lexical associations and their subject content, particularly when prototypical expressions that align with authentic reading examples are identified (Flowerdew, 2015). This employment of authentic textbook material related to their studies situates learners in the context of their academic genre demands (Ismayilli Karakoç et al., 2022).

Different activities can be proposed for novice university students to explore online by reflecting on key word use over specific texts. This approach can be combined with beyond-the-text activities, such as discussions about concepts being addressed, so that lexical retention and conceptual understanding are targeted (Curado Fuentes, 2024b; Tavares Pinto, 2024). Collaborative and repetitive tasks involving key vocabulary have been shown to result in more sustained linguistic knowledge (Itawa et al., 2024).

### III. RESEARCH QUESTIONS

Three research questions guide the design of this case study:

- 1) What is the effect of BDDL on ESP courses with mixed linguistic proficiencies?
- 2) How can BDDL be integrated for lexical and text comprehension exploitation along these ESP courses?
- 3) Does BDDL constitute a positive approach in this type of ESP contexts?

### IV. METHODOLOGY

Two groups of ESP learners using BDDL were examined by relying on a sequential mixed-methods design, based on Teddlie and Tashakkori (2009) and Grove and Cipher (2024), keeping track of learning traits and developments through class activities, pre-, mid-, and post-tests, in-class polls, and end-of-semester interviews.

#### IV.1. Participants

The students were enrolled in Specific English I, a one-semester compulsory subject offered during the first year of the double degree of Business Management and Tourism at University of Extremadura. The overall number of participants was 26: 14 in the Spring 2023 semester (average age = 18.1 years; 8 female, 6 male), and 12 in 2024 (average age = 18.2 years; 5 female, 7 male). Most learners possessed intermediate linguistic levels in English, but others had higher or lower levels. An online English

proficiency test ([cambridgeenglish.org/test-your-english/general-english/](https://cambridgeenglish.org/test-your-english/general-english/)), taken in class at the beginning of the semesters, determined that in 2023, six students were at B1 (intermediate level), three at A2 (low-intermediate), two at B2 (upper-intermediate), two at C1 (advanced), and one at A1 (low/basic). In 2024, six students were at B1, two at B2, two at A2, and two at C1.

Students' academic backgrounds and preliminary views on English learning were also examined at the beginning of each semester. A simple questionnaire with multiple-choice items was answered online. Overall, the responses revealed a preference for oral skills (speaking and listening), with more students prioritizing these skills in 2023 (85.3 percent) and fewer in 2024 (67.7 percent), when more learners (33.3 percent) chose writing as their priority. Additionally, most students in 2023 reported that their secondary school English subjects had emphasized the study of grammar, whereas the proportion decreased (41.6 percent) in 2024, with half of the class referring to the main use of listening/speaking approaches in previous years. In contrast, group work was their chief preference for classroom dynamics (78.5 percent in 2023 and 83.3 in 2024). In 2024, due to the expansion of GenAI in academic contexts, some questions were added about the use of GenAI tools. The result was that almost everyone (91.5 percent) admitted to using either ChatGPT or Gemini for academic work, and 58.3 percent used GenAI for English practice, mainly, for translation (five students) and lexical / grammatical consultation (three participants).

## **IV.2. Course design**

The main organizational difference between the two courses was the duration and frequency of the BDDL developments. In 2023, BDDL was confined to a four-week period (16 class hours). In contrast, in 2024, BDDL was practiced regularly every two to three days throughout the semester, totalling approximately 20 hours. Both courses utilized similar tools, including Versatext, Just the Word, Corpus Mate, SKELL, NetSpeak, YouGlish, Reverso, WordReference, and the Collins dictionary. In 2024, ChatGPT and

Gemini were added.<sup>1</sup> Other online resources like Kahoot, Acadly, and Quizlet were employed to explore and review topic-specific vocabulary and short text readings.<sup>2</sup>

Table 1 provides a sequenced outline of activities and their integration into the courses, along with the instruments used for the measurement and evaluation of learning developments. Additionally, some changes made in 2024 are indicated. Most BDDL activities were conducted before, during, or after exploring topics by selecting academic texts from freely accessible repositories (e.g., open textbooks for Social Sciences at <https://open.umn.edu/opentextbooks/subjects/social-sciences>).

**Table 1.** *BDDL in the two courses*

BDDL in both courses	Instruments	Only in 2024
Introduction: keywords and texts (Acadly discussions)	Pre-test (Google form) on vocabulary and text comprehension	No pre-test was taken before the BDDL sessions
Questions answered in pairs / groups using cloud keywords (Versatext)	Poll utility (in Acadly) to collect students' impressions with the tools and activities after sessions	A greater number of polls was conducted
Texts provided by the instructor; questions answered in pairs / groups (Versatext, Reverso, online dictionaries)	Mid-test (Google form) on vocabulary and text comprehension	This was the first test taken by the 2024 group
Activities with texts selected by students in groups (Versatext)	Acadly discussions: Students' impressions with overall tools and activities	Students worked on texts in mixed-linguistic proficiency groups

1 All the tools were used freely online: <https://versatext.versatile.pub>, <http://www.just-the-word.com>, <https://corpusmate.com>, <https://skell.sketchengine.eu/#home?lang=en>, <https://netspeak.org>, <https://youglis.com>, <https://www.reverso.net>, <https://www.wordreference.com>, <https://www.collinsdictionary.com/dictionary/english>, <https://chatgpt.com>, <https://gemini.google.com>.

2 Acadly, Kahoot, and Quizlet were utilized in their fully licensed versions, funded by the University of Extremadura. Acadly (<https://www.acadly.com>) served as a platform for course management and online tracking of students' activities, including quizzes and polls. Kahoot (<https://kahoot.com>) facilitated interactive games designed to challenge learners with essential concepts and vocabulary, while Quizlet (<https://quizlet.com>) was employed to engage students with key linguistic content from the course. These three tools provide a gamified approach to learning, consistently receiving high ratings in end-of-course surveys over the years.

BDDL in both courses	Instruments	Only in 2024
Activities with diverse tools (Corpus Mate, Just the Word, SKELL, NetSpeak, YouGlish)	Post-test (Google form) on vocabulary and text comprehension	GenAI tools were also used
	Students' oral presentations and impressions about the tasks	Individual interviews were conducted with four students

The activities under “BDDL in both courses” (Table 1) were the following:

- 1) As introductory exercises, students were asked to identify keywords in digital texts without using any tools. The concept of keywords was explained as topic-based (Scott 2010), and students worked in groups to identify them and explain why they were keywords, associating them to main text themes.
- 2) Learners were guided along the analysis of cloud keywords extracted from digital texts. This activity involved students' use of the Word Cloud and Concordance utilities in Versatext with the same previous texts from the introductory sessions to compare concordances and sentence examples.
- 3) Different academic texts were provided by the instructor, with subsequent activities based on the instructor's previous analyses. These tasks explored dynamics such as decoding specific keywords in context, inferring concrete meaning from linguistic use over concordances, identifying part-of-speech words in different expressions, and analysing meta-textual references (see examples in Curado Fuentes, 2024a). Online dictionaries and translators were also used to answer specific questions on linguistic meaning.

Students relied on keywords for linguistic comprehension in both courses. Most initial questions asked by learners involved the specification of keywords to be used. For example, the text comprehension question “What is the highest cost listed in the estimated budget for the project infrastructure?” led them to scrutinize different key nouns in the Word Cloud (e.g., “budget”) to find the actual answer, scanning concordances. They also proceeded similarly with keywords for reversed translation activities (e.g., finding and decoding the expression *factores clave para captar clientes* using the keyword “target,” which they had to analyse as a verb, not a noun).

4) Students were free to choose and analyse their own texts in subsequent activities, following the instructor's guidelines. Most texts and activities were shared on Padlet (<https://padlet.com>), an easy-to-use board allowing students to smoothly navigate between resources and to copy and paste content as needed. In 2024, each group included at least one higher-linguistic proficiency student.

In general, both courses identified specific vocabulary and analysed lexical occurrences from their generated keyword clouds. Most texts used by learners were about sociology and tourism management. In these texts, students noticed words collocating with keywords, and they examined them as words that co-occurred with these keywords over concordances. Students also used the Profiler utility in Versatext to explain whether these keywords were general, academic, or text-specific. In most cases, they chose keywords classified as academic (e.g., "sample" and "factors" in a text on sociology for tourism). In various cases, learners had to cope with part-of-speech disambiguation, as in, for example, the keyword "advertising," often seen as an adjective, but also working as a noun (e.g., "form of advertising" and "role of advertising").

5) Different online tools were exploited for analytical comparison. For example, with SKELL, students worked with keyword sketches, such as verbs associated with a key noun from previous texts (e.g., "innovation"), or words occurring as subjects of previously seen verbs (e.g., "manage"). In Corpus Mate, concordances were chiefly explored to compare lexical meaning and use across different academic texts (e.g., a previous keyword, "corporate," in Law, Economics, and History). This contrastive analysis was aimed at enhancing lexical knowledge and expanding textual comprehension with keywords and concordances.

In 2024, learners took on GenAI tools for this complementation of BDDL tasks. Unlike concordancers, these tools were already familiar to students. The activities with GenAI were intended as reviewing activities. For example, learners had to ask about their keyword-based findings (Table 2), with steps 1 and 2 instructing learners to compare and discuss key linguistic content, and with step 3 requesting for the application of linguistic knowledge to writing (see an example of an A2 student's answers in Table 2).

**Table 2.** Example of GenAI activity

Questions	Answers (A2 student)
Ask ChatGPT to provide five keywords based on your text. Compare these with your own keywords from the previous activity.	<i>Sociological, analysis, culture, innovation, creativity</i> <i>Two are the same and three different: Issues, market, tourism</i>
Ask ChatGPT to use the keywords in different contexts related to business. Compare these sentences with your own from the previous activity. How are they different?	<i>A company that understands the sociological impact of environmentalism may promote eco-friendly products</i> <i>The sociological trends of tourism are in my analysis and not here</i>
Write three full sentences of your own with any keywords from your list. Use at least three keywords and three words that combine with them.	<i>I think that innovation is key for the future of tourism</i> <i>The market-driven approach is better</i> <i>When companies innovate their products are improve</i>

Along the courses, some discussions with students unfolded online (using Acadly) and orally in class. Students’ impressions with various class tasks were shared. Finally, oral presentations, assigned in groups, were aimed at evaluating learners’ use of linguistic and content information derived from BDDL activities.

### IV.3. Evaluation instruments

As shown in the “Instruments” column (Table 1), a pre-test was administered at the start of the 2023 module, but it was omitted in 2024. Instead, this pre-test was taken as a mid-term test a few weeks into the semester of 2024 (see “Only in 2024” in Table 1). This change was made to compare learners’ test performance after already using BDDL, feasible in 2024 due to the longer duration of the BDDL activities during this semester.

These tests consisted of 15 questions: 10 multiple-choice items focusing on academic vocabulary, and five reading comprehension questions. The vocabulary part included tasks such as identifying synonyms and collocations with keywords, English/Spanish equivalents, and concepts with key terms. The reading comprehension section required short written answers based on textbook extracts (see test examples in

Curado Fuentes, 2024a). Students were told that these tests would not influence their final grades in the course so that they could perform unrestrained by this pressure and to the best of their knowledge. The online completion of the tests (pre-, mid-, post-) led to 42 scores (3 tests x 14 students) in 2023, and 24 (2 x 12) in 2024. All test scores showed normal distributions according to the Shapiro-Wilk test (see Table 3 below).

Additionally, to gather learners' impressions, the poll utility included in Acadly (see "Instruments" in Table 1) was used at the end of some class sessions. These polls were aimed at rating learners' levels of agreement with different statements about the BDDL tools and activities on a 5-point Likert scale. In 2023, students responded to three 10-item polls, whereas in 2024, a total of six 10-item polls were completed. For the final survey analysis, the different poll items were summarized into a set of 13 items. For example, various poll items rating participants' perceptions about the usefulness of different activities for vocabulary learning were summarized as a final item named "usefulness of BDDL for vocabulary" (see Table 6 in Results). Each student's set of scores was averaged for the final survey analysis.

Three key ideas were targeted in these polls: 1) Perceived usefulness of the activities for linguistic learning (usefulness), 2) difficulty with tools (usage), and 3) potential use of these mechanics in the future (usability), based on Hua et al. (2024). Students' scores were analysed using Excel and online tools (Wessa 2024). Shapiro-Wilk tests were also conducted to evaluate whether these scores were normally distributed, and the results were negative. Therefore, all the survey responses were treated as non-normally distributed (Table 3).

**Table 3.** Shapiro-Wilk test results on the score distributions

Instruments	2023	2024
Pre-tests	W(14) = .97, $p = .933$	
Mid-tests	W(14) = .91, $p = .184$	W(12) = .94, $p = .531$
Post-tests	W(14) = .91, $p = .138$	W(12) = .94, $p = .540$
Polls	W(14) = .83, $p = .011$	W(12) = .9, $p < .001$

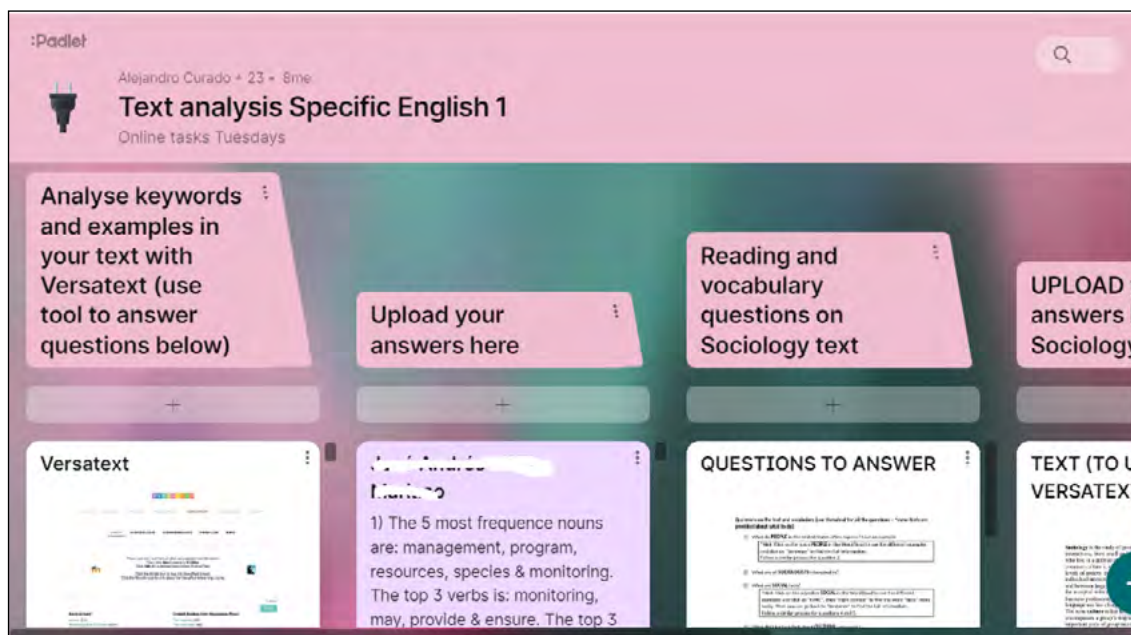
Additionally, four students were interviewed by the instructor in 2024 (see “Only in 2024” in Table 1). These semi-structured interviews relied on a consistent set of questions for each interviewee, allowing for more detailed explorations based on their responses. These four students were chosen because they represented the four main linguistic proficiencies: C1, B2, B1, and A2. The goal was to compare their insights with previous polls and classroom feedback, aiming at capturing students’ personal opinions, expressed in Spanish to allow for greater depth of expression. The interviews lasted between five and 10 minutes per participant. Their responses were recorded and transcribed using the TurboScribe tool (<https://turboscribe.ai/es>). The author of this study read the transcriptions several times to identify key ideas and themes, combining a deductive analysis with an inductive approach. This method is effective with qualitative data of a manageable size (Azungah, 2018).

## **V. RESULTS**

### **V.1. Classroom observations and discussions**

Students generally preferred the use of smartphones over laptops for online interactivity. However, for text selection and analysis, they relied on laptops more, sharing devices in groups, especially in 2024.

Overall, students in both courses felt eager to select their own topics in groups, and they stated that this mechanics allowed for greater freedom in the analytical tasks. This linguistic probing generally contributed towards enhancing various instances of metalinguistic awareness, since they shared questions and answers, which they posted on Padlet (Figure 1) for all to see.



**Figure 1.** Screenshot of text analytical activities with their own texts

The more linguistically proficient learners within each group often guided the exploration of texts, suggesting key terms to explore. However, participatory contributions were progressively made by lower linguistic-level learners. Online and oral discussions ensued from this sharing of activities. In general, higher-linguistic proficiency students stated more meta-linguistic ideas, especially positive reviews of the activities for lexical gain. Lower-level learners stated more concerns with linguistic complexity in the texts, especially over concordances.

In the GenAI activities, linguistically proficient students also provided more sophisticated answers, interacting with the tool by deploying more prompts and linguistic nuances. For example, they understood and compared different phrasal and lexical options rendered by the tool, choosing fitting options. In turn, lower-level students interacted less with the tools, producing simpler answers. The positive side was that these linguistic findings were compared by all learners with their own previous linguistic discoveries. As a result, answers were critically examined, not just copy-pasted.

Students also did well on the oral presentation assignments, based on the content explored with BDDL. This task was done as project work, for which they organized a topic (e.g., innovation in tourism) by introducing the topic, outlining main parts or sections

according to keywords, explaining main aspects, and concluding with an emphasized or strong viewpoint. The connection with content analysed in previous BDDL activities was generally realized by learners, as they used different keywords and expressions in their deliveries, thus reinforcing metacognition and linguistic awareness.

## V.2. Test results

Paired-sample t-tests were conducted to compare linguistic knowledge and text comprehension performance before, during, and after the interventions: pre-, mid-, and post-test scores for 2023, and mid-term tests and post-test scores for 2024. The results revealed significant improvements in all cases, with p-values below 0.05 for all comparisons. In 2023, the post-test scores were compared with mid-test scores, and mid-test scores with pre-test scores (Table 4).

**Table 4.** Significance values for test scores within groups

Year	Mid-test at $p < .05$	Post-test at $p < .05$
2023	.00002569	.03295
2024		.000003983

All the test scores were also compared with students' linguistic proficiencies to examine whether individual scores correlated with their language competences. Table 5 demonstrates a strong positive correlation in all cases, with values exceeding the critical values of Pearson's R at 0.45 (N = 14-2) in 2023 and 0.49 (N = 12-2) in 2024. These findings suggest a proportional relationship between general English proficiency and test performance, indicating a clear tendency for changes in linguistic proficiency to correspond with changes in test performance in the same direction. In both years, however, this strong prediction decreases in the post-tests. The reason is that lower-proficiency learners' scores progressively rose, even more so in 2024.

**Table 5:** Correlation values of test scores with linguistic proficiencies

Year	Pearson correlation coefficient value
2023 pre-test	0.856919
2023 mid-test	0.815243
2023 post-test	0.730646
2024 mid-term test	0.782701
2024 post-test	0.668856

### V.3. Poll results

Students' perceptions were measured by averaged poll scores. These scores were tested for internal consistency and reliability within each of the three categories (see Table 6), and the result was that Cronbach's Alpha values exceeded 0.8 in all categories, except for the second one (usage), which reported a value of 0.705. As a result, this final 13-item set was considered valid for analysis.

Due to the small number of students in each course, Table 6 gathers mean scores from both years combined (i.e., 26 students). These scores were calculated and classified according to three variables: Learners' linguistic proficiency (with letter C referring to C1 and B2-level students, B to B1, and A to A2 and A1), learning approach preferences (O = oral, W = written), and classroom dynamics (I = individual, G = group). This classification was based on learners' scores in the general English level tests and on their introductory questionnaire preferences at the beginning of the semesters (see Participants in Methodology).

**Table 6.** Mean scores for categories of poll items according to group variables

CATEGORY	SURVEY ITEMS	LINGUISTIC PROFICIENCY	LEARNING APPROACH	CLASS DYNAMICS
Usefulness (4 items)	BDDL was useful for...:	C = 3.48		
	Vocabulary learning	B = 3.06	O = 3.43	I = 3.11
	Reading skills	A = 1.99	W = 2.59	G = 2.95
	Grammar			
Usage (4 items)	BDDL was easy to use in terms of...:	C = 3.85		
	Concordances	B = 3.26	O = 3.49	I = 3.55
	Keywords	A = 2.76	W = 2.79	G = 3.41
	Text management			
Usability (5 items)	In the future, BDDL can be used for...:			
	Self-learning	C = 3.66		
	Other English subjects	B = 3.05	O = 3.21	I = 3.31
	Other university subjects and projects	A = 2.79	W = 2.81	G = 3.35
	Academic reading			
Academic writing				

Higher values corresponded to higher-proficiency (C) learners and to students who prefer oral tasks (O). Lower-level (A) students' scores were lower by comparison with C and B students, especially in terms of their perceived usefulness of BDDL for linguistic learning (underlined in Table 6); however, these A students' scores increased regarding usage and usability, especially in 2024. In particular, the scores for tool navigability and applicability to other English subjects rose more in 2024, with their consideration of GenAI usage as either easy or very easy notably contributing to this increase. Overall, the items receiving higher scores in both years were vocabulary knowledge (for usefulness), keywords (usage), and other English subjects (usability).

To identify whether these mean scores were significantly different from each other, Kruskal-Wallis H-tests were run for the three-group measurement (i.e., linguistic

proficiency), and Mann-Whitney U-tests for the two-group reports (i.e., learning approach and class dynamics). P values (at  $p < .05$ ) were used to ascertain any meaningful differences (indicated with asterisks in Table 7).

**Table 7.** P-values from non-parametric tests according to group variables. \*Significant differences

CATEGORY	LINGUISTIC PROFICIENCY	LEARNING APPROACH	CLASS DYNAMICS
Usefulness	.00044*	.00697*	.08372
Usage	.00193*	.00775*	.09563
Usability	.00199*	.05678	.09849

Based on these statistical findings, the only variable without any significant score differences was class dynamics, suggesting that learners' preferences for either group-based or individual work did not influence their appreciation of BDDL. In contrast, for the other two variables, significant relationships were found. So, post-hoc Dunn's tests were run using a Bonferroni corrected alpha value of 0.017 to specify the groups of students' scores which significantly differed. Table 8 shows the groups with significant score variations in each category.

**Table 8.** Student groups' significantly different scores in each category

CATEGORY	LINGUISTIC PROFICIENCY	LEARNING APPROACH
Usefulness	A	W
Usage	C, B, A	W
Usability	C, B, A	--

Therefore, in terms of statistical significance, it is confirmed that lower linguistic-proficiency students perceived less usefulness in BDDL for language learning. On the other hand, advanced learners appraised BDDL as easier and more likely to be used in the future than intermediate and low-intermediate students, whereas intermediate students assigned proportionally higher values than low-intermediate participants.

In turn, students favouring written approaches considered BDDL less relevant for linguistic learning and less easy to use, whereas these students' perceptions did not significantly differ from oral-skills-focused participants' regarding their appraisal of BDDL for future use.

#### **V.4. Interviews**

The four interviewees in 2024 were asked similar questions, with variations tailored to explore opinions and perceptions according to this outline:

a: General question:

*What do you think about these BDDL activities and tools?*

b: Question about linguistic knowledge:

*What linguistic aspects do you think you improved?*

*What do you think about the use of these utilities for language learning?*

*What were some positive and negative things?*

c: Procedural/navigational question:

*How did you start working with the tools?*

*How did you find out the answers?*

*What were your main difficulties?*

*Which utilities became more convenient for exploring texts and vocabulary?*

Their opinions and ideas were grouped according to themes. Some examples were the following (with my own translation of their comments):

1) Learning approach:

*a: I liked using all the tools for clarifying linguistic doubts with them (C1 student)*

*b: I liked the use of the tools in groups because we could post the information together and learn from others (A2 student)*

2) Linguistic competence:

*c: The advantage was that we could go over the expressions in groups, and this helped me to understand the expressions (...) which were sometimes complicated (A2)*

*d: Versatext was really good for vocabulary, (...) and then, we could choose different examples to translate the ones needed (...) Google translator helped us a lot too (B1)*

3) Problem-solving:

*e: We sometimes made mistakes because the examples (in the concordancer) were not clear for some answers. (...) We found more examples and finally managed to answer and translate the meaning correctly (B1)*

*f: I had problems because I didn't understand (the activity) (...) and all of my peers knew more vocabulary and grammar than me. (...) Yet, with the help of my classmates I could find out how to do it (A2)*

4) GenAI:

*g: I have really liked ChatGPT, (...) it helped us a lot for the answers with vocabulary (B1)*

*h: We used ChatGPT too, and I liked it for this objective (...) [but] I don't trust ChatGPT very much (...) because it is repetitive (B2)*

5) Academic English:

*i: These resources have provided me with important insight into the key vocabulary for my own studies because I often find them (the terms) in English readings (for other subjects) (C1)*

*j: The texts are important for our English in business (...) This type of activity is useful and I liked the Versatext activities (...) I enjoyed using it on my own and yes, I think it would be useful for English (B2)*

The students generally expressed positive views regarding BDDL, highlighting Versatext and GenAI for productive task development and linguistic intake. The advanced learners commented on the actual potential of these resources and strategies for future study

in ESP / EAP. In contrast, the low-intermediate student reflected more on the difficulty of facing complex language tasks, and yet, she realized the productive use of working in groups to solve problems and of the activities for linguistic knowledge enhancement. The importance of these activities for lexical competence was mentioned more by the intermediate student, who also appreciated combining online translators and ChatGPT.

## **VI. DISCUSSION**

Based on learners' ideas and performance, some key aspects can be discussed about the application of BDDL in mixed linguistic-proficiency scenarios of ESP.

Students with higher linguistic proficiencies perform better in tasks and tests. Linguistic competence correlates with linguistic performance. Advanced learners explicitly relate BDDL affordances to the enhancement of linguistic learning and academic English, appreciating BDDL advantages more, just like experts may value innovative developments in a field more positively than non-experts. However, intermediate-level learners also realize the usefulness of BDDL for language learning, and they perceive the advantageous utility and usability of the tools more distinctively than lower-level learners. In turn, the significant score improvements from one test to the next are achieved by all learners, which demonstrates that all types of learners increasingly improve their lexical knowledge and text decoding skills. In the case of low-competence participants, their scores progressively improve even more in 2024, suggesting that a longer exposure to BDDL increasingly benefits linguistic knowledge at low-intermediate linguistic levels.

Combining various tools positively influences learning performance and attitudes. Learners' successful management of utilities generally leads to task acceptance and positive reviews. Students' cognitive and linguistic skills also develop constructively by using GenAI as a complementary tool for linguistic probing. Their realization that these tools add to linguistic knowledge works in favour of integrating GenAI within BDDL.

BDDL fosters metacognitive and metalinguistic awareness among learners. Linguistic reflections occur along the activities, with advanced learners often leading comments and discussions. In turn, intermediate students contribute thoughts on their recognition of linguistic benefits, whereas lower-linguistic level participants express more modest appreciations. Two primary reasons seem to arise: lexical difficulty and academic English analysis. Even though linguistic / text analysis with the tools tends to become easy and profitable, lower-linguistic level learners find more difficulty in terms of linguistic decoding for meaning discovery. However, they showcase improved analytical skills in combining simple query tools (e.g., concordances and GenAI) with group work and peer collaboration, as observed in class and pointed out in discussions and interviews. High degrees of interaction and consultation with classmates tend to transfer into positive emotional effects for learners regardless of class dynamics preferences.

The use of concordances is also more effective with BDDL. Some studies have examined that more complaints and difficulties are related to concordances in low-intermediate ESP contexts (e.g., Curado Fuentes, 2015). However, in this case study, concordance developments are integrated better within task mechanics, as these utilities consistently receive higher scores in the surveys (compared to other years). Simpler query interfaces and text-based tools seem to work as a main influence on such positive reactions.

Finally, some noteworthy variations seem to arise in terms of learning preferences. However, if based solely on polls, with such a small sample of students, these observations should be made cautiously. Perhaps, more distinctively, less favourable aspects, such as lower scores given by students who prefer writing skills, should be noted, indicating that this type of BDDL activities are not appropriate for academic writing. Additionally, most advanced and upper-intermediate linguistic-proficiency learners chose oral skills as their preference, which coincides with higher scores assigned to the benefits of BDDL. These students thus seem more likely to integrate BDDL into future academic work.

## **VII. CONCLUSIONS**

This case study has provided some insights into how learners can adapt to BDDL for linguistic developments in ESP.

Answering the first research question of this case study, it is confirmed that BDDL can be used productively in mixed linguistic-proficiency ESP contexts. Students' performances and impressions demonstrate this observation. Even if lower-linguistically proficient students face greater challenges with linguistic probing, the combination of simple query tools and dynamic group work tends to motivate them, mitigating misgivings about this type of approach. This finding aligns with previous results on the importance of collaborative work for low-linguistic level learners using DDL (e.g., Lusta et al., 2023).

In relation to the second research question, integrating BDDL in ESP courses may be done in different ways, but one key element seems to be the implementation of simple text query applications so that BDDL can engage lower-linguistic level learners more suitably, combining other tools to enrich linguistic learning. A key point is task repetition and long exposure to BDDL with different tools. Clear directions and scaffolding along the way are important to ease learners' cognitive efforts throughout the activities, in agreement with some DDL studies (e.g., Timmis, 2015). In sum, different types of activities may be designed, but a key aspect is that learners experience the advantages afforded by the flexibility, ease of use, and group-aided developments of BDDL.

Finally, to answer the third research question, this case study confirms that BDDL constitutes a positive approach in ESP settings. BDDL is realized as a useful strategy to introduce learners into vocabulary and texts within their fields, since learners recognise and appreciate its usefulness for their academic context. This realization creates an opportune space for their practice with higher-order skills, such as field-specific content comprehension, analytical skills, and problem-solving. This type of focus is considered important at this university stage. However, this case study presents notable limitations regarding student sample size and longitudinal observations (i.e.,

tracking the same students over the years); therefore, expanding the study in these directions can significantly strengthen the findings and observations about the impact of BDDL on ESP.

## REFERENCES

- Arce, Norman Patrick Harvey, & Cuadros Valdivia, Ana Maria (2020). Adapting competitiveness and gamification to a digital platform for foreign language learning. *International Journal of Emerging Technologies in Learning (IJET)*, 15(20), 194-209. <https://doi.org/10.3991/ijet.v15i20.16135>
- Argyroulis, Vassilis (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>
- Azungah, Theophilus (2018). Qualitative research: deductive and inductive approaches to data analysis. *Qualitative Research Journal*, 18(4), 383-400. <https://doi.org/10.1108/QRJ-D-18-00035>
- Bárcena, Elena, Read Timothy, & Arús, Jorge (Eds.). (2015). *Languages for specific purposes in the digital era*. Springer.
- Basturkmen, Helen (2010). *Developing courses in English for specific purposes*. Palgrave Macmillan.
- Belcher, Diane (2009). What ESP is and can be: An introduction. In Diane Belcher (Ed.), *English for specific purposes in theory and practice* (pp. 1-19). Michigan ELT.
- Boulton, Alex (2012). Hands-on/hands-off: Alternative approaches to data-driven learning. In James Thomas & Alex Boulton (Eds.), *Input, process and product: Developments in teaching and language corpora* (pp. 152-168). Masaryk University Press.
- Boulton, Alex (2016). Integrating corpus tools and techniques in ESP courses. *Asp*, 69, 113-137. <https://doi.org/10.4000/asp.4826>

- Boulton, Alex (2024). Corpus linguistics and research syntheses: Respect for data in making sense of text and context. *Asp*, 86, 71-82. <https://doi.org/10.4000/12ry1>
- Boulton, Alex, & Cobb, Tom (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 1–46. DOI:10.1111/lang.12224.
- Boulton, Alex, & Vyatkina, Nina (2024). Expanding methodological approaches in DDL research. *TESOL Quarterly*, 58(3), 1193-1204. DOI: 10.1002/tesq.3269
- Boulton, Alex, & Pérez-Paredes, Pascual (2025). Data-driven learning: Pedagogy and technology. In Regine Hampel & Ursula Stickler (Eds.), *Bloomsbury handbook of language learning and technology*. Bloomsbury.
- Carter, Ronald, & McCarthy, Michael (1995). Grammar and the spoken language. *Applied Linguistics*, 16(2), 141–158. doi:10.1093/applin/16.2.141
- Chambers, Angela (2022). What is data-driven learning? In Anne O`Keeffe & Michael McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 416-429). Routledge.
- Charles, Maggie (2012). ‘Proper vocabulary and juicy collocations’: EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes*, 31, 93–102. <https://doi.org/10.1016/j.esp.2011.12.003>
- Charles, Maggie (2014). Getting the corpus habit: EAP students’ long-term use of personal corpora. *English for Specific Purposes*, 35, 30–40. DOI : 10.1016/j.esp.2013.11.004
- Cobb, Tom (2018). From corpus to CALL. The use of technology in teaching and learning formulaic language. In Anna Sivanova-Chanturia & Ana Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 192-211). Taylor & Francis.
- Criollo-C, Santiago, González-Rodríguez, Mario, Guerrero-Arias, Andrea, Urquiza-Aguiar, Luis F., & Luján-Mora, Sergio (2024). A Review of Emerging Technologies and Their Acceptance in Higher Education. *Education Sciences*, 14(1), 10. <https://doi.org/10.3390/educsci14010010>

- Crosthwaite, Peter & Baisa, Vit (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast! *Applied Corpus Linguistics*, 3(3). <https://doi.org/10.1016/j.acorp.2023.100066>
- Curado Fuentes, Alejandro (2015). Exploiting keywords in a DDL approach to the comprehension of news texts by lower-level learners. In Agnieszka Lenko-Szymanska & Alex Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 177-198). John Benjamins.
- Curado Fuentes, Alejandro (2023). Corpus affordances in foreign language reading comprehension. In Kieran Harrington & Patricia Ronan (Eds.), *Demystifying corpus linguistics for English language teaching* (pp. 99-118). Palgrave MacMillan.
- Curado Fuentes, Alejandro (2024a). Digital reading in English for business management and tourism: The importance of micro-skills for the curriculum. *REALIA. Research in Education and learning innovation archives*, 33, 16-34. DOI:10.7203/realia.33.27586
- Curado Fuentes, Alejandro (2024b). Reconsidering Pérez-Paredes' informal learning contexts of DDL. In Peter Crosthwaite (Ed.), *Corpora for language learning: Bridging the research-practice divide* (pp. 225-226). Routledge.
- Flowerdew, John, & Peacock, Matthew (2001). *Research perspectives on English for academic purposes*. Cambridge University Press.
- Flowerdew, Lynn (2009). Applying corpus linguistics to pedagogy. *International Journal of Corpus Linguistics*, 14(3), 393–417. doi:10.1075/ijcl.14.3.05flo
- Flowerdew, Lynn (2015). Corpus-based research and pedagogy in EAP: From lexis to genre. *Language Teaching*, 48(1), 99-116. DOI:10.1017/S0261444813000037
- Grove, Susan, & Cipher, Daisha (2024). *Statistics for nursing research*. Elsevier Health Sciences.
- Hadley, Gregory, & Hadley, Hiromi (2021). Exploring the impact of data-driven learning in extensive reading. In Pascual Pérez-Paredes & Geraldine Mark (Eds.), *Beyond concordance lines: Corpora in language education* (pp. 149-176). John Benjamins.

- Hadley, Gregory (2025). Reading and data-driven learning. In Lee McCallum, & Dara Tafazoli (Eds.), *The Palgrave encyclopedia of computer-assisted language learning*. Palgrave Macmillan.
- Huan, Yuan Fang, Lu, Xiaofei, & Guo, Qian (2024). Independent corpus consultation for collocation use in academic writing by L2 graduate students. *System*, 127, 103515. <https://doi.org/10.1016/j.system.2024.103515>
- Ismayilli Karakoç, Aynur, Ruegg, Rachel, & Gu, Peter (2022). Beyond comprehension: Reading requirements in first-year undergraduate courses. *Journal of English for Academic Purposes*, 55. <https://doi.org/10.1016/j.jeap.2021.101071>.
- Iwata, Akira, Kanayama, Kohei, & Kasahara, Kiwamu (2024). Effects of pair work and individual work on paired-associate vocabulary learning in an EFL context. *International Journal of Applied Linguistics*. <https://doi.org/10.1111/ijal.12610>
- Johns, Tim (1991). Should you be persuaded: Two examples of data-driven learning. In Tim Johns & Philip King (Eds.), *Classroom concordancing* (pp. 1-16). Centre of English language studies, University of Birmingham.
- Karakose, Turgut, & Tülübaş, Tijen (2023). How can ChatGPT facilitate teaching and learning: Implications for contemporary Education. *Educational Process: International Journal*, 12(4), 7-16. DOI: 10.22521/edupij.2023.124.1
- Laufer, Batia, & Ravenhorst-Kalovski, Geke C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15-30.
- Laufer, Batia, & Cobb, Tom (2019). How much knowledge of derived words is needed for reading? *Applied Linguistics*, 41(6), 1–29. <https://doi.org/10.1093/applin/amz051>
- Lee, Sanghee, Kim, Sunghee, & Jung, Yang-soo (2019). Effects of DDL approach on English productive and receptive vocabulary knowledge and reading ability. *The Journal of Studies in Language*, 35(3), 341-360. DOI:10.18627/jslg.35.3.201911.341

- Liu, Guangxiang Leon, Darwin, Ron, & Ma, Chaojun (2024). Exploring AI-mediated informal digital learning of English (AI-IDLE): a mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer-Assisted Language Learning*, 1-29. <https://doi.org/10.1080/09588221.2024.2310288>
- Law, Locky (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, 6, 100174. <https://doi.org/10.1016/j.caeo.2024.100174>
- Luo, Shuqiong, & Zou, Di (2024). University learners' readiness for ChatGPT-assisted English learning: Scale development and validation. *European Journal of Education*, 0, e12886. <https://doi.org/10.1111/ejed.12886>
- Lusta, Amel, Demirel, Özcan, & Mohammadzadeh, Behbood (2023). Language corpus and data-driven learning (DDL) in language classrooms: A systematic review. *Heliyon*, 9, e22731. DOI:10.1016/j.heliyon.2023.e22731
- Nel, Carisma (2008). Learning style and good language learners. In Carol Griffiths (Ed.), *Lessons from good language learners* (pp. 49–60). Cambridge University Press.
- Nishizawa, Hitoshi, Yoshioka, Takayoshi, & Ichikawa, Yuri (2018). Book-talk: An activity to motivate learners to read autonomously in a foreign language. *Journal of Language and Cultural Education*, 6(1), 145–157. <https://doi.org/10.2478/jolace-2018-0010>
- Ordoñana, Guillamón Carlos, Pérez-Paredes, Pascual, & Aguado-Jiménez, Pilar (2024). Pedagogic natural language processing resources for L2 education: Teachers' perceptions and beliefs. *Language Value*, 17(2), 60-99. DOI: <https://www.doi.org/10.6035/languagev.8522>
- Pan, Mengru, Lai, Chun, & Guo, Kai (2024). Effects of GenAI empowered interactive support on university EFL students' self-regulated strategy use and engagement in reading. *The Internet and Higher Education*. <https://doi.org/10.1016/j.iheduc.2024.100991>

- Paribakht, Sima, & Webb, Stuart (2016). The relationship between academic vocabulary coverage and scores on a standardized English proficiency test. *Journal of English for Academic Purposes*, 21, 121-132. <https://doi.org/10.1016/j.jeap.2015.05.009>
- Pérez-Paredes, Pascual, Ordoñana, Carlos, & Aguado-Jiménez, Pilar (2018). Language teachers' perceptions on the use of OER language processing technologies in MALL. *Computer Assisted Language Learning*, 31(5–6), 522–545. <https://doi.org/10.1080/09588221.2017.1418754>
- Pérez-Paredes, Pascual (2024). Data-driven learning in informal contexts? Embracing Broad Data-driven learning (BDDL) research. In Peter Crosthwaite (Ed.), *Corpora for language learning: Bridging the research-practice divide* (pp. 211-221). Routledge.
- Philip, Gill (2010). Classroom concordancing in the 21st century: The new generation. *Language Forum*, 36(2), 1-20.
- Ronan, Patricia (2023). Learning to teach EFL with corpus linguistics approaches: A survey of teacher training students' attitudes. In Keiran Harrington, & Patricia Ronan (Eds.), *Demystifying corpus linguistics for English language teaching* (pp. 19-38). Palgrave MacMillan.
- Scott, Mike (2010). Problems in investigating keyness, or clearing the undergrowth and marking out trails. In Marina Bondi & Mike Scott (Eds), *Keyness in texts* (pp. 43–58). John Benjamins.
- Song, Tianjiao, & Reynolds, Barry Lee (2022). The effects of lexical coverage and topic familiarity on the comprehension of L2 expository texts. *TESOL Quarterly*, 56(2), 763–774. <https://doi.org/10.1002/tesq.3100>
- Tate, Tamara, & Warschauer, Mark (2022). Digital tools for promoting social reading. In Mark Brooke, Shenfa Zhu, Anuradha Ramanujan, Tetyana Smotrova & Sylvia Sim (Eds.), *Selected papers from the 6th CELC symposium for English language teachers* (pp. 4-17). Centre for English Language Communication.

Timmis, Ivor (2015). *Corpus linguistics for ELT: Research and practice*. Routledge.

Tavares Pinto, Paula (2024). Using a data-driven learning approach to teach English for Academic Purposes and improve higher education internationalisation. In Peter Crosthwaite (Ed.), *Corpora for language learning: Bridging the research-practice divide* (pp. 227-239). Routledge.

Teddlie, Charles, & Tashakkori, Abbas (2009). *Foundations of mixed methods research: integrating quantitative and qualitative approaches in the social and behavioral sciences*. Sage Pubs.

Wessa, Patrick (2024). *Free statistics software*. Office for research development and Education. Version 1.2.1. <https://www.wessa.net>

Xie, Qing (2024). Individual differences and non-English majors' engagement in business English language classrooms in the Chinese university context. *Ibérica*, 47, 149-174. <https://doi.org/10.17398/2340-2784.47.149>

Xu, Manfei, Chen, Xiao, Liu, Xiaobin, Lin, Xiaoyue, & Zhou, Qiaoxin (2019). Using corpus-aided data-driven learning to improve Chinese EFL learners' analytical reading ability. In Simon K.S. Cheung, Jianli Jiao, Lap-Kei Lee, Xuebo Zhang, Kam Cheong Li & Zehui Zhan (Eds.), *Technology in education. Pedagogical innovations* (pp. 15-26). Springer.

*Received: 29 April 2025*

*Accepted: 09 July 2025*