

Designing Collaborative Resources in Data-Driven Learning to Accommodate Learner Diversity

Muhammad Rudy¹ , Emi Emilia¹ , Wawan Gunawan¹  and Destiani² 

¹ English Language Education, Universitas Pendidikan Indonesia, Indonesia

² Universitas Pendidikan Indonesia, Indonesia and Universitas Lampung, Indonesia

Abstract

Data-Driven Learning (DDL) usually relies on existing corpora like British National Corpus (BNC) and Corpus of Contemporary American English (COCA), but these may not align with students' needs when writing an abstract. To address this issue, designing and implementing a tailored corpus could be a solution. This study explores students' perceptions during the collaborative creation of an abstract corpus (RQ1) and its strengths and weaknesses in an academic writing process (RQ2). This research was conducted as a case study involving 80 out of 94 final-year engineering students who completed the task of building a corpus of abstracts to serve as a writing reference. Despite varying interests, the participants collaborated to compile a corpus of relevant abstracts. They collaboratively collected, annotated, and analyzed the corpus. To support corpus development and the DDL process, Google Drive was used as online storage, and AntConc software was used as an analysis tool. The data were collected through interviews and video recordings. The video recordings were used to capture the DDL process starting from preparation to corpus utilization. The retrospective interviews with five randomly selected participants were conducted to explore participants' perceptions. The findings suggest that collaborative corpus development enhances students'

E-mail: muhammadrudy@upi.edu, emi.emilia@upi.edu, wagoen@upi.edu & destiani@fkip.unila.ac.id

understanding of abstract writing genres. Different research interests did not hinder the corpus development and utilization. An observable weakness was the corpus tool installation problem. In conclusion, the collaborative corpus development fostered inclusivity and deeper involvement, highlighting the potential of DDL to meet diverse students' needs.

Keywords: Collaborative, corpus development, data-driven learning, interdisciplinary projects

Introduction

Interdisciplinary projects have been long implemented in engineering higher education, including in telecommunication engineering, to bring solutions in real life (Rahman et al., 2025; van den Beemt et al., 2020). As a result, the final projects in telecommunication engineering education, known as Capstone Projects, have been applied in many sectors like smart, safe, and sustainable road infrastructure, agriculture, education, etc. (Rahman et al., 2025; Yao et al., 2023). However, the critics of the interdisciplinary projects emerged with the question of a learning program that should accommodate various specific projects that have specific genres and needs (Markauskaite et al., 2024). The obstacles in the English as Foreign Language (EFL) context regarding English for academic writing are that the lecturer must accommodate various specific needs related to topics (Yang, 2022). This is a challenge in order to facilitate all students' various needs.

DDL has become an influential approach in language learning, particularly for developing writing competence through corpus-based exploration (Johns, 1991; Pérez-Paredes & Boulton, 2025). Traditionally, DDL relies on large, pre-compiled corpora such as British Academic Writing English (BAWE) used by Crosthwaite et al. (2023) and Michigan Corpus of Upper-Level Student Papers (MICUSP) used by Karpenko-Seccombe (2024) or even built-in corpora in online analysis tools like SketchEngine used by Crosthwaite and Steeples (2022). Those corpora provide learners with access to authentic language input and usage patterns (Boulton & Cobb, 2017). However, generic corpora often fall short in meeting the specialized needs of learners engaged in discipline-specific writing (Egan, 2019; Flowerdew & Petrić, 2024). For instance, in this study context, engineering students writing abstracts must adhere to distinct rhetorical conventions and terminological expectations that are rarely represented in general corpora. Prior research has acknowledged this issue, noting that learners struggle when linguistic models do not align with their field-specific tasks (Flowerdew, 2015).

One proposed solution is the development of a specialized corpus, which offers both pedagogical and disciplinary relevance (Charles, 2012; Esfandiari & Allaf-Akbary, 2024; Sfeir, 2025). Collaborative corpus building, where students collect, annotate, and analyze texts, provides opportunities to engage more deeply with language while simultaneously constructing a useful writing resource. Charles (2012) and Esfandiari and Allaf-Akbary (2024) argued that such learner-led initiatives support genre awareness and foster autonomy. When students become co-constructors of linguistic resources, the process itself, like annotating and data cleaning, can enhance their genre awareness and engagement (Flowerdew & Petrić, 2024; Pérez-Paredes, & Boulton, 2025; Sfeir, 2025) in ways that static corpora cannot.

In this context, the study investigates how collaborative corpus development can support abstract writing among 80 final-year engineering students who conducted interdisciplinary research projects. It is different from the study conducted by Charles (2012), where the 47 multidisciplinary students work individually to build a corpus for personal use. It is also not similar to the study conducted by Sfeir (2025), where the 11 multidisciplinary students worked either individually or collectively to build their corpora. Thus, this study tried to investigate the students' perception (RQ1) and explore the strengths and weaknesses (RQ2) of the collaborative corpus building in abstract writing.

Methodology

Research Design

This study employed a qualitative case study design (Kumar, 2011) to explore how collaborative corpus supports abstract writing. The population consisted of 94 final-year engineering students enrolled in a scientific writing course at a private university in West Java. The study was conducted in the odd semester in 2024, taking three out of sixteen meetings of a compulsory course, namely Capstone Project. The timeline of DDL implementation can be seen in Figure 1, which is elaborated in the following subsections.

Participants and Corpus Development

94 students from the sixth and seventh semesters were grouped by the Telecommunications Engineering Department of the university. They took a 4-credit Capstone Course where the focus was writing a capstone report. The course lasted for 16 meetings in a semester. As part of the course completion and report submission, abstract writing training is conducted in the last three meetings. The participants were grouped based on their research interests

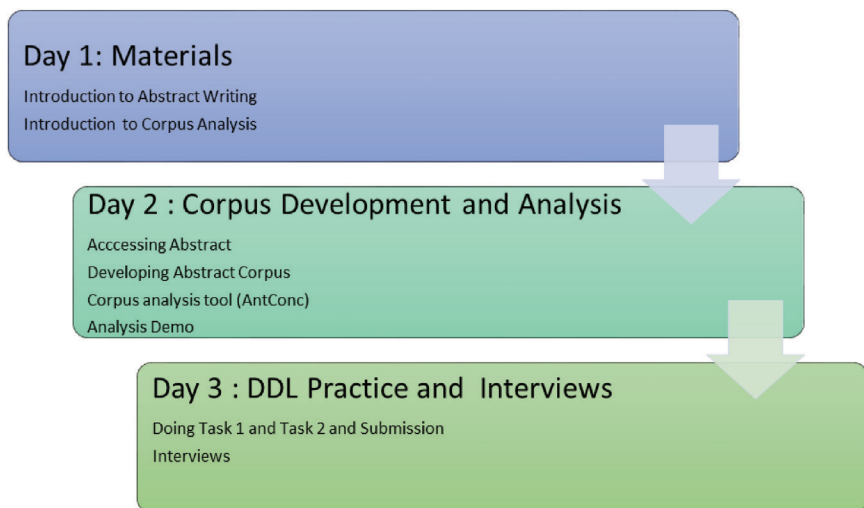


Figure 1. Teaching procedure

(see Appendix A), varying from machine learning to anxiety detection. Each group consisted of two to four people.

The development of the abstract corpus was completed by 80 students from 28 project groups. Each project group contributed at least two abstracts, which were compiled and stored using a folder in Google Drive. Each member of the group was instructed to collect abstracts from journals relevant to their capstone projects. The corpus was not treated with data cleaning, known as dirty corpus (Charles, 2012), to reduce unwanted elements in the text like equations (Sfeir, 2025). The students were instructed to annotate the abstracts manually for identifying abstract sections such as Introduction, Methods, Results, and Discussions (IMRD). This activity aimed to raise their awareness of genre structure and prepare them for corpus analysis.

Teaching Procedure

The DDL teaching procedure shown in Figure 1 was conducted in the thirteenth meeting. The first meeting was conducted to ask consent from the participants and give an explanation of the study. The materials explained on this day were an introduction to abstract writing and corpus analysis. The second meeting was conducted to train corpus development and analysis demo.

Once the corpus was ready, in the last meeting, the students used AntConc (Anthony, 2024) to analyze lexical and phrasal patterns based on abstract

sections. Specific attention was given to word lists and frequently used expressions—particularly phrases following IMRD sections (Swales, 1990). The students worked on answering two questions: “What are ten important words for your abstract?” and “What five writing expressions will you use to start writing the abstract introduction, method, result, and discussion sections?” The participants submitted their answers together with the justification for why they chose those ten words and five phrases. The answers were collected in the Google Form.

Instruments and Analysis

In order to address the two research questions, students’ perception and strengths and weaknesses, this study employed documentation and an interview. The documentation included two submitted tasks, photographs, and videos. The submitted tasks were scrutinized to identify reasons the participants chose words and phrases. The photos and videos captured group discussions and peer correction during the teaching process.

To gain deeper insight into students’ experiences, the interviewers announced the recruitment of interviewees at the end of the abstract writing program. An open-ended interview was conducted with five self-selected participants or volunteers (Brinkmann & Kvale, 2018). The interview focused on students’ perceptions of the collaborative process, corpus usability, and the challenges they encountered. The member checking process was also done by referring to the volunteers’ responses upon the two submitted tasks in the interview process.

The analysis stages were conducted by following suggestions from Kumar (2011, p 278): theme identification, coding, and classification. These three stages were conducted by the second author. The theme identification was through reading the Google form response about selected words and phrases as well as the justifications. The justifications tended to fall on two streams which are, genre awareness and engagement/ownership (Flowerdew & Petrić, 2024; Pérez-Paredes, & Boulton, 2025; Sfeir, 2025). The code was given for the reasons or the causes of choosing the words or phrases. The coded response then was classified into themes (Braun & Clarke, 2006). Peer checking towards the themes and codes was conducted by the third and fourth authors. The coding results can be seen in Table 1 and Table 2.

Similar to the task documents, the interviews were transcribed, and then the second author coded and identified themes. The third and fourth authors conducted peer checking on the transcription, codes, and themes. The improper transcription and code differences were discussed to reach agreement for correction.

Table 1. Distribution of task 1 justifications

No	Code	Number	Percentage %
1	Capstone Project Relevance	28	24.35
2	Abstract Structure	19	16.52
3	Linguistic Function	19	16.52
4	Field Relevance	15	13.04
5	Frequency	12	10.43
6	Academic Needs	10	8.70
7	Conceptual Representation	8	6.96
8	Ease of Understanding	4	3.48
	TOTAL	115	100

Table 2. Distribution of task 2 justifications

No	Code	Number	Percentage %
1	Conceptual representation	47	41.59
2	Abstract structure	26	23.01
3	Ease of understanding	13	11.50
4	Frequency	10	8.85
5	Academic needs	9	7.96
6	Linguistic function	3	2.65
7	Field relevance	3	2.65
8	Capstone Project relevance	2	1.77
	TOTAL	113	100

Results and Discussions

Task 1 showed that most students chose the ten important words that related to their capstone project, with 28 responses (24.35%). For example, participant number 73 showed in Figure 2 that the participant chose ten important words: *fiber, network, IoT, learning, optical, machine, detection, communication, monitoring, and networks*. He said those words were chosen because they were relevant to his Capstone Project. He said, “because these words are common vocabulary in telecommunications engineering, which are often used in capstone design topics in this department.”

In contrast, Task 2 was dominated by Conceptual Representation Code with 41.59%. In choosing five phrases to begin writing each section of IMRD,

participant 33 chose five phrases: *In this paper*, *To overcome the problem*, *This study compares*, *Based on these problems*, and *The method used* as shown by Figure 3. He said, “These five phrases were chosen because they are relevant to conveying research results effectively, both to state findings, prove hypotheses, and describe the impact of research in a professional manner.”

Regarding the RQs in this study, the answers are presented in the following subsections. The perceptions and plan to use the corpus for learning are elaborated. The strengths and weaknesses of corpus-based language learning

	E	F	G
1	app	10 kata penting	Mengapa 10 kata tersebut penting?
73	516521	fiber,network,iot,learning ,optical,machine,detection,communication,monitoring,networks	karena kata kata tersebut merupakan kosa kata umum di lingkungan teknik telekomunikasi yang sering digunakan dalam suatu topik capstone design di jurusan ini. with : karena ini kata hubung yang sangat berperan system : karena ini menjelaskan secara singkat method : langkah dalam penjelasan tugas yang performance : tugas yang sudah dikerjakan past technology : technology pada umumnya inti sari
74	773100	with: svstem; method; per	

Figure 2. Task 1 reasons coded with capstone project relevance.

	E	F	G	H
1		Di bagian mana Anda mengerjakan?	5 Frasa penting	Mengapa 5 frasa tersebut penting?
32	5214	Results	2. The findings of 3. Simulation result show 4. Based on 5. In this	Frasa-frasa tersebut merupakan kalimat pembuka untuk menjelaskan hasil dari penelitian
33	3255	Method	In this paper, to overcome the problem, this study compares, based on these problems, the method used	Karena 5 frasa tersebut menunjukkan bagian method pada abstrak paper atau jurnal suatu proyek dengan spesifik The Methodology: karena memberikan gambaran umum tentang pendekatan yang digunakan dalam penelitian.

Figure 3. Task 2 reasons coded with conceptual representation code.

in the study context highlight genre awareness and engagement. Lastly, the participants reveal a negligible weakness during DDL in academic writing class.

Student Perceptions of Collaborative Corpus (RQ1)

The perception of the participants after completing DDL in the abstract writing program was positive. The corpus-based language learning was relatively easy to follow and it was useful and helpful for their writing process. The five interviewees agreed that the process was easy as can be seen in the excerpts (the complete version in Appendix B).

- Participant 1: Ais
It's easy because, the user interface is also very pleasant
- Participant 2: Aan
It's really easy to find frequently occurring words.
- Participant 3: Kar
Interviewer: So, there are benefits, right? Then, were there any difficulties while learning?
Interviewee: Not yet.
Interviewer: No difficulties yet. So, it's easy.
- Participant 4: Hai
It's easy, sir, because it's very easy to search for words, and that's what the corpus showed. It makes things easier for us.
- Participant 5: Nur
I think I understand because the corpus tool is also quite easy to understand.

All four interviewees explicitly stated that DDL in the writing process was easy. Only one interviewee, Participant 3, did not mention it. However, Participant 3 said during the process he did not find any difficulties, which implied it was easy. Furthermore, the signs of positive perception on DDL for abstract writing intensified as the participants claimed the DDL was useful and helpful. These claims have been proven by the following excerpts:

- Participant 1: Ais
It's quite helpful, so I can better understand the content of abstracts, and also what abstract writing patterns are like.
- Participant 2: Aan
Yes, it is useful, sir. Well, it seems like, it's easier to read without having to click one by one first, we'll read what the results are like, sir.

Participant 4: Hai

I'll probably use it, because it was very helpful.

Participant 5: Nur

It's very useful, especially when searching for words and their frequency.

In addition, the positive perception was also stated by the interviewees. They realized that they acquired new knowledge after the DDL practice in writing abstract. The excerpts are:

Participant 3: Kar

My impression is that I have gained new knowledge for writing

Participant 4: Hai

It broadened my knowledge and helped me learn something new.

The participants' positive perception towards DDL for writing abstracts is because it is easy to practice, useful, helpful, and knowledge-enriching. This encouraging finding is in line with the study conducted by Charles (2012) that her students were enthusiastic about using DIY corpus for writing because it was useful to improve their writing. Crosthwaite and Steeples (2022), though they used large pre-compiled corpora for assisting secondary students in writing, also have found that the students expressed positive perceptions of DDL.

Positive perception was also found by Spivey (2023) when he studied DDL implementation with elementary students. He scaffolded the students by using two types of DDL, paper-based and computer-based. He said that the perception possibly changes over time, and students preferred paper-based DDL after facing difficulties practicing computer-based. The participants in the current study seems to have a consistent perception since they are from telecommunication engineering, where working with computers is part of their daily routine.

In line with a recent study conducted by Esfandiari and Allaf-Akbary (2024), the perception of students (83.7%) taught with hands-on DDL for identifying interactional metadiscourse in EFL writing was positive. DDL motivated the students to realize interactional metadiscourse since it activated autonomy and self-discovery. It is because they were given a chance to develop a corpus and analyze it by using AntConc software, similar to this study. This indicated that a specialized corpus built collaboratively for DDL brings positive perception among participants.

Besides positive perception, collaborative DIY DDL brought strengths and weaknesses discussed in the following subsection.

Strengths and Weaknesses in Writing Practice (RQ2)

Genre Awareness and Engagement

From Tasks 1 and 2, the coded response directed us to draw two streams; Genre Awareness and Engagement. Table 3 informs that Genre Awareness in Task 1 is derived from abstract structure, linguistic function, and academic needs. It occurs 48 times (41.74%) across the participants’ responses.

In addition, the Task 2 presented in Table 4 revealed that the Genre Awareness, similar with the first task, is indicated by the abstract structure, linguistic function, and academic needs. It occurs 38 times (33.63%). The reasonings stated by the participants after choosing ten words and five phrases

Table 3. Theme and code from task 1

No	Themes	Codes	Number	Percentages %
1	Engagement	Capstone Project Relevance, Concept Representation, Field Relationship, Frequency, Ease of Understanding	67	58.26
2	Genre Awareness	Abstract Structure, Linguistic Function, Academic Needs	48	41.74
	Total		115	100

Table 4. Theme and code from task 2

No	Themes	Codes	Number	Percentages %
1	Engagement	Capstone Project Relevance, Concept Representation, Field Relationship, Frequency, Ease of Understanding	75	66.37
2	Genre Awareness	Abstract Structure, Linguistic Function, Academic Needs	38	33.63
	Total		113	100

brought to the conclusion that the collaborative DIY corpus in the DDL enabled them to realize the pattern and genre in abstract.

Across interviews, participants expressed increased genre awareness, particularly in identifying the structure and function of academic abstracts. Three participants noted:

Participant 1: Ais

It has been quite helpful, so I can better understand **the content of abstracts** and also **the patterns of how abstracts** are written.

Participant 2: Nur

From there, **several phrases were displayed**, and then when writing an abstract, we get an idea like, Oh, **the opening words** when writing, especially in English, should be something like this.

Participant 3: Aan

There was no difficulty at all. Because when I looked at the papers that were collected, I thought, Oh, so in these papers, there must be an **abstract, introduction, and results**. Then, within the results, there must be **analysis and discussion**.

Those three participants reflected a heightened sensitivity to abstract writing convention. This hands-on engagement with genre conventions appears to have bridged the gap between theoretical knowledge and practical application.

Another significant finding was the enhanced engagement that resulted from students' sense of ownership over the corpus-building process. Rather than relying on pre-existing materials, students curated texts that reflected their research interests, fostering a sense of personal investment. Data indicated that the students reported feeling more joyful in this task compared to traditional writing tasks (as stated by Participant 2 and 3 above). The collaborative nature of the project—along with its relevance to students' ongoing thesis work—made the activity both meaningful and contextually grounded, reinforcing previous findings that learner-led DDL promotes autonomy and engagement.

Table 3 and 4 have shown that during Task 1 and 2, the students' engagement was seen in the codes like Capstone Project Relevance, Concept Representation, Field Relationship, Frequency, and Ease of Understanding. The response, like in Figure 2, indicated that the participants were strongly bound to their individual capstone projects, although during the DDL implementation, they together built a single corpus.

The interviews conducted with the five participants brought evidence of engagement by using a collaborative DIY corpus. The participants exhibited three out of five engagement codes: Capstone Project Relevant, Concept Representation, and Frequency. Below are the excerpts demonstrating the reasons why they chose a word or a phrase during finishing Task 1 and 2.

- Participant 1: Ais [Frequency] (Commenting about Task 2)
 Interviewer: You chose *the results, the experiment, and the model*. Why did you choose them?
 Interviewee: Because they have a higher frequency, there are several words that are suitable.
- Participant 2: Aan [Capstone Project Relevant] (Commenting about Task 1)
 Interviewer: Why did you choose *performance*?
 Interviewee: Well, because in our results, we display *performance*, that relates to *accuracy* and *precision*, sir. In the paper we wrote, that's *performance*.
 Interviewer: This is closely related to the paper your group wrote? So, your research results show *performance*.
 Interviewee: Yes, that's right.
- Participant 3: Kar [Concept Representation] (Commenting about Task 1)
 Interviewer: You chose *the, and, of, are, results, keyword, research, title, and detection*. Why are the words *the, and, and of* important?
 Interviewee: Because usually *the* is at the beginning, sir. If you don't use *the*. It feels like something is missing in the abstract.
- Participant 4: Hai [Frequency] (Commenting about Task 1)
 Interviewer: Why did you choose those words?
 Interviewee: They appear a lot, sir.
- Participant 5: Nur [Concept Representation] (Commenting about Task 2)
 Interviewer: Why did you choose the phrase *the purpose*?
 Interviewee: Why did I choose that? Because when we create research, there must be a purpose. It must be explained in the abstract. If there is a *the purpose*, there must be a goal for producing that abstract.

A stronger indication of engagement in completing the task during DDL is shown by videos. A Video 1 and Video 2 captured that the students were busy with their laptops to complete Task 1 and 2. Little discussions among the

participants were captured. Even though in the video they were discussing in their group, each of the members submitted his personal answers based on his own preferences. Their discussion was about the stages of submitting the task in order to avoid mistakes.

After compiling the findings from coded responses, videos, and interviews, engagement is relevant to the studies conducted by Charles (2012) and Esfandiari and Allaf-Akbary (2024). Both previous studies discussed the utilization of DIY corpus for writing teaching. Charles (2012) let her students use an individual corpus to help them writing. Esfandiari and Allaf-Akbary (2024) used a corpus created from 20 research articles from two reputable journals to be used in DDL practice to assess interactional metadiscourse. Both pieces of research agreed that DIY corpus improved learners' engagement during the DDL. The difference in the recent study is that the corpus was built collaboratively among interdisciplinary participants as shown by Task 1 and 2, video recording, and interviews.

The strengths of DDL by using a collaborative DIY corpus in this study were facilitating genre awareness and fostering learning engagement during the learning process. The collaborative DIY corpus in this study successfully accommodated the inclusivity of participants' research areas by providing various options of words and phrases to choose from, yet the participants kept focusing on their topic of interest. The corpus reflected a wide range of topics, providing diverse language use samples and discipline-relevant strategies. This variety ensured that each student could find examples closely aligned with their abstract writing needs.

Drawback

Behind the strengths of using a collaborative DIY corpus in DDL previously mentioned, a notable challenge also emerged. First, some students stated difficulty in installing the corpus analysis tool, AntConc, from the website due to internet connection crowding (this part was not recorded). During the DDL practice, all participants accessed the Anthony Laurence website at the same time. The download process was taking time which hindered the smoothness of learning because after the download process the participants were required to install it onto their personal laptop. The DIY corpus by using a similar analysis tool, may face this problem. The experience of this occurred in the interview with Participant 4.

Participant 4: Hai

The only **issue we faced was maybe just during installation.**

The internet connection problem experienced by the participants seems not to relate directly to the practice of DDL with a collaborative corpus. This is probably due to the participants being from a telecommunication engineering major where computer software and program are their core competency. Moreover, positive perception of the participants shown previously confirms that there was no significant difficulty during DDL practice.

Conclusions

This study examined how collaborative corpus development supports abstract writing among telecommunication engineering students within a DDL framework. The participants had a positive perception of the DDL practice. DDL seems easy, useful, and helpful in assisting with the writing of an abstract. This study also indicates that the genre awareness and engagement of the participants were enriched. The participants were involved in curating and analyzing a corpus relevant to their academic interests, enhancing genre awareness and engagement. The process of DDL with a collaborative DIY corpus encouraged participants in understanding disciplinary writing conventions through student-selected texts. The weakness of a collaborative corpus in DDL was unclear due to familiarity bias for telecommunication engineering students.

The limitation of this study was the short time of the implementation of the collaborative corpus in DDL. This becomes a gap to fill by the future research. Although the participants have interdisciplinary topics with different needs, they are from a telecommunications engineering major, which is familiar with the tool, so that they did not face any difficulties on DDL. Future research should involve more diverse participants to strengthen the inclusivity of DDL.

Acknowledgement

We would like to thank the editor and reviewers for the invaluable suggestions and criticisms that helped improve this manuscript.

Funding

This work was supported by Indonesian Education Scholarship (Beasiswa Pendidikan Indonesia/**BPI**), Center for Higher Education Funding and Assessment (Pusat Pembiayaan dan Asesmen Pendidikan Tinggi/**PPAPT**) Kemdiktisaintek, and Indonesian Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan/**LPDP**)

References

- Anthony, L. (2024). AntConc (Version 4.3.1) [Computer Software]. Waseda University. <https://www.laurenceanthony.net/software/AntConc>
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. <https://doi.org/10.1111/lang.12224>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brinkmann, S., & Kvale, S. (2018). *Doing Interviews*: SAGE Publications. (The Sage Qualitative Research Kit).
- Charles, M. (2012). Proper nouns: Pedagogical tools for teaching academic writing. *Journal of English for Academic Purposes*, 11(1), 54–66. <https://doi.org/10.1016/j.jeap.2011.11.005>
- Crosthwaite, P., Luciana, & Wijaya, D. (2023). Exploring language teachers' lesson planning for corpus-based language teaching: a focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning*, 36(7), 1392–1420. <https://doi.org/10.1080/09588221.2021.1995001>
- Crosthwaite, P., & Steeples, B. (2022). Data-driven learning with younger learners: exploring corpus-assisted development of the passive voice for science writing with female secondary school students. *Computer Assisted Language Learning*, 37(5–6), 1166–1197. <https://doi.org/10.1080/09588221.2022.2068615>
- Egan, T. (2019). Non-representativeness in corpora: Perils, pitfalls and challenges. *CogniTextes*. Article 1772. <https://doi.org/10.4000/cognitextes.1772>
- Esfandiari, R., & Allaf-Akbary, O. (2024). Assessing interactional metadiscourse in EFL writing through intelligent data-driven learning: the Microsoft Copilot in the spotlight. *Language Testing in Asia*, 14(1). <https://doi.org/10.1186/s40468-024-00326-9>
- Flowerdew, L. (2015). Corpus-based research and pedagogy in EAP: From lexis to genre. *Language Teaching*, 48(2), 153–174. <https://doi.org/10.1017/S0261444814000019>
- Flowerdew, L., & Petrić, B. (2024). A critical review of corpus-based pedagogic perspectives on thesis writing: Specificity revisited. *English for Specific Purposes*, 76, 1–13. <https://doi.org/10.1016/j.esp.2024.05.003>
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *English Language Research Journal*, 4, 1–16.
- Karpenko-Secombe, T. (2024). *Practical use of corpora in teaching argumentation. In Corpora for language learning* (pp. 111–130). Routledge. <http://dx.doi.org/10.4324/9781003413301-9>

- Kumar, R. (2011). *Research Methodology: a step-by-step guide for beginners*. SAGE Publications Ltd.
- Lee, D., & Swales, J. M. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56–75. <https://doi.org/10.1016/j.esp.2005.02.010>
- Lin, M. H. (2021). Effects of Data-Driven Learning on College Students of Different Grammar Proficiencies: A Preliminary Empirical Assessment in EFL Classes. *SAGE Open*, 11(3). <https://doi.org/10.1177/21582440211029936>
- Markauskaite, L., Schwarz, B., Damşa, C., & Muukkonen, H. (2024). Beyond disciplinary engagement: Researching the ecologies of interdisciplinary learning. *Journal of the Learning Sciences*, 33(2), 213–241. <https://doi.org/10.1080/10508406.2024.2354151>
- Pérez-Paredes, P., & Boulton, A. (2025). *Data-driven Learning in and out of the Language Classroom*. Cambridge University Press. <https://doi.org/10.1017/9781009511384>
- Rahman, M. ., Sarwar, H. ., Tin, T. T. ., & Kader, M. A. . (2025). A systematic literature review on final year capstone project in an engineering curriculum. *Edelweiss Applied Science and Technology*, 9(6), 1408–1440. <https://doi.org/10.55214/25768484.v9i6.8146>
- Sfeir, M. (2025). Do-it-yourself corpora and discipline-specific writing: A focus on the benefits of corpus building and cleaning (in the age of large language models). *English for Specific Purposes*, 79, 70–86. <https://doi.org/10.1016/j.esp.2025.04.002>.
- Spivey, M. (2023). Data-driven learning and young learners: Perceptions and attitudes of Japanese elementary school students toward EFL corpus-based activities. *JALT CALL Journal*, 19(3), 394–416. <https://doi.org/10.29140/jaltcall.v19n3.1014>
- van den Beemt, A., MacLeod, M., & van der Veen, J. (2020). *Interdisciplinarity in tomorrow's engineering education*. In J. van der Veen, N. van Hattum-Janssen, H.-M. Jarvinen, T. de Laet, & I. Ten Dam (Eds.), *Engaging Engineering Education: Book of Abstracts SEFI 48th Annual Conference* (pp. 61-74). European Society for Engineering Education (SEFI). <https://www.sefi.be/wp-content/uploads/2020/11/Proceedings-2020-web.pdf>
- Yang, L. (2022). Focus and Interaction in Writing Conferences for EFL Writers. *SAGE Open*, 12(1). <https://doi.org/10.1177/21582440211058200>
- Yao, H., Xu, Z., Hou, Y., Dong, Q., Liu, P., Ye, Z., ... & Wang, D. (2023). Advanced industrial informatics towards smart, safe and sustainable roads: A state of the art. *Journal of traffic and transportation engineering* (English edition), 10(2), 143-158. <https://doi.org/10.1016/j.jtte.2023.02.001>

Appendix A

Students Research Interests

No	Topics	Group Member (Persons)
1	SDN Based 5g Core Network Emulator System with Graphical Interface for 5g Core Network Design	3
2	Observation of the Effect of Anxiety Conditions on the Shift in the Working Frequency of EEG Signals	4
3	Machine Learning Based Multi Disturbance Prediction Modeling for Disturbance Handling in Dense Wavelength Division Multiplexing (DWDM) Longhaul	3
4	Multi Disturbance Prediction Modeling Based on Machine Learning for Handling Disturbances in Optical Network Cables	3
5	Design of Earthquake Detection System Based on Internet of Things	4
6	Resource Allocation in Device to Device Communication Using SCMA Model for 5G Network	3
7	Implementation of Open-Source Cloud-Based WiFi Controller to Improve User Experience in an Open Library Case Study	4
8	Resource Block Allocation for Hetnet (Heterogeneous Network) in 4G AND 5G	4
9	Resource Allocation in Hybrid Interweave-Underlay Scheme in Cognitive Radio Network	3
10	WIH (Monitoring of Water Quality and Automatic Feeding of Tilapia Fish Ponds Based on IoT)	3
11	HRS (Sentiment Analysis of Tourist Attractions in Bandung)	3
12	UMY (Development of Mobile Applications Based on Machine Learning and Internet of Things for Environmental Monitoring Maggot Cultivation as Organic Waste Decomposers	4
13	HRS (Sentiment Analysis of Tourist Attractions)	2
14	Implementation of Internet of Things (IoT) in Aquaponic System for Automatic Monitoring and Controlling of Fish Pond and Hydroponic Maintenance	4

(Continued)

No	Topics	Group Member (Persons)
15	Use of Artificial Intelligence in Audio Watermarking to Improve Information Hiding Performance	4
16	Estimation of Bit Error Rate in Free Space Optic Using Machine Learning	4
17	FlexiWAN-Based uCPE Device as SD-WAN Vendor Lock-In Solution	2
18	Design and Realization of RT RW NET In Girimukti Village	3
19	Trading-dong 2.0: Forex Expert Advisor using Strategy Based on Price Action and Order Block Under The Control of Artificial Intelligence	4
20	Design of Hybrid Access Network for Unifiber Provider by Asianet From Ciwastra Point of Presence to Batununggal Area Bandung	4
21	Automatization of Diagnosis and Claims Process Drugs for Optimization of BPJS Claims by Hermina Hospital Arcamanik	3
22	Optimization of Student Counseling Services at a University Through Digital Applications with NLP Technology	3
23	Improving Network Service Quality Through Ad Blocking and DNS Filtering	3
24	Gender Identification Based on Bite Marks	4
25	Design of Repeated Bending and Twist Test Tools for Patchcord and Precon Fiber Optic Cables at a Test House	4
26	Temperature and Humidity Monitoring and Control System for Incubator for Tempe Fermentation Process Based on IOT Technology	3
27	Design and Analysis of Single Frequency Network on WHOOSH KCIC Between Padalarang Station and Tegalluar Summarecon Station Lines	3
28	Android Application Readers of Javanese Script In Javanese Manuscript and Literary Works	3
		94

Appendix B

Excerpt of Perception

Participant 1: Ais

- Interviewer: I am conducting research related to the learning activities that have been carried out. I have a few questions. First, how do you think learning using the corpus was?
- Interviewee: in my opinion, because this is my first time too. I know... what's the name of this website, because this software... ...it's quite helpful, so I can better understand the content of abstracts, and also what abstract writing patterns are like.
- Interviewer: More or less, there are benefits, right?. In your opinion, was it difficult or easy to learn using this tool?
- Interviewee: Easy.
- Interviewer: Easy?
- Interviewee: It's easy because, the user interface is also very pleasant, sir.
- Interviewer: Okay, Then, after that, are you going to use AntConc or another corpus tool to help you write abstracts?
- Interviewee: I'll try using it again.
- Interviewer: So, will you use it in the future?
- Interviewee: Yes.

Participant 2: Aan

- Interviewer: you've already learned how to use corpus. What's your impression?
- Interviewee: Corpus is that application, right? It's really easy to find frequently occurring words, sir. Because we just input all of them (abstract files). Then we can search for the results, or the frequently occurring words, or how many files are included (Range). Previously, we usually searched for keywords in the abstracts. Previously, to find the most frequently occurring ones (words) was difficult, sir. So, with this application, it's easier, sir.
- Interviewer: So, do you want to continue using this method?
- Interviewee: If I want to write another paper or journal, maybe I will. But I don't know yet, I don't have any plans to do that (write a paper) again, sir.

- Interviewer: So, do you think it's useful?
- Interviewee: Yes, it is useful, sir. Well, it seems like, it's easier to read without having to click one by one first, we'll read what the results are like, sir.
- Interviewer: KWIC interface?.
- Interviewee: yes, that's easier, sir

Participant 3: Kar

- Interviewer: Well, this is related to writing the abstract class that you have been involved; what are your impressions?
- Interviewee: My impression is that I have gained new knowledge for writing.
- Interviewer: Oh, new knowledge. New knowledge in what way?
- Interviewee: Previously, I was confused about writing. But the software makes it easier.
- Interviewer: So, there are benefits, right? Then, were there any difficulties while learning?
- Interviewee: Not yet.
- Interviewer: No difficulties yet. So, it's an easy.

Participant 4: Hai

- Interviewer: According to you, is learning like that, using a corpus tool, difficult or easy?
- Interviewee: It's easy, sir, because it's very easy to search for words, and that's what the corpus showed. It makes things easier for us, sir.
- Interviewer: Were there any difficulties?
- Interviewee: No, sir. The only problem was the installation process, right?
- Interviewer: Oh yes. Was the learning helpful?
- Interviewee: Very helpful, sir. It broadened my knowledge and helped me learn something new.
- Interviewer: Okay, okay. Will you use corpus again when writing abstracts?
- Interviewee: For me, I'll probably use it, because it was very helpful.

Participant 5: Nur

- Interviewer: Okay, you've attended the class, especially the one about writing abstracts. This can be done using a corpus. Do you understand or master the steps taught previously?
- Interviewee: I think I understand because the corpus tool is also quite easy to understand.

- Interviewer: Oh, it's quite easy. Also, there weren't any difficulties. So, do you think this kind of learning, using a corpus, is useful?
- Interviewee: Of course, it's very useful, especially when searching for words and their frequency. Using regular tools takes a long time, but using a corpus is fast, right?
- Interviewer: Oh, right, okay. And also... Do you think what you've learned will be useful for writing abstracts?
- Interviewee: Yes, it's useful.
- Interviewer: What kind of example?
- Interviewee: For example, when creating phrases, for example, we searched for phrases using the corpus (tool). From there, there are several phrases that have been displayed, and then to make the abstract we have an idea for, oh, the initial words in writing the abstract, especially in English, are like this.

