



## DATA-DRIVEN LEARNING IN PRE-TERTIARY EFL EDUCATION: A CRITICAL REVIEW OF APPROACHES, OUTCOMES, AND UNRESOLVED ISSUES

Muhammad Andrie Bagia<sup>1</sup>, Dwita Laksmi Rachmawati<sup>2</sup>

Universitas Muhammadiyah Gresik, Indonesia<sup>1,2</sup>

e-mail: [andriebagia@gmail.com](mailto:andriebagia@gmail.com)

Diterima: 04/05/2026; Direvisi: 09/05/2026; Diterbitkan: 18/05/2026

### ABSTRAK

Pembelajaran bahasa Inggris sebagai bahasa asing (EFL) pada tingkat pra-perguruan tinggi masih didominasi oleh pendekatan berbasis buku teks dan pembelajaran yang berpusat pada guru, sehingga membatasi akses peserta didik terhadap penggunaan bahasa autentik. Sebagai alternatif, Data-Driven Learning (DDL) menawarkan pendekatan berbasis korpus yang memungkinkan peserta didik mengeksplorasi pola bahasa melalui data autentik. Penelitian ini bertujuan untuk mensintesis pendekatan pembelajaran DDL, mengidentifikasi hasil belajar yang dilaporkan, serta menganalisis tantangan implementasi DDL dalam konteks EFL pra-perguruan tinggi. Penelitian ini menggunakan pendekatan tinjauan literatur analitis kualitatif dengan melakukan pencarian dan seleksi artikel secara sistematis melalui Google Scholar dan sumber akademik terkait berdasarkan prinsip PRISMA. Data dianalisis menggunakan sintesis tematik untuk mengidentifikasi pola utama terkait pendekatan pembelajaran, hasil belajar, dan kendala implementasi. Hasil penelitian menunjukkan bahwa DDL umumnya diterapkan melalui pendekatan teacher-mediated dan guided discovery yang menyeimbangkan kemandirian belajar dengan scaffolding guru. DDL secara konsisten mendukung perkembangan kosakata, kesadaran kolokasi, dan language noticing, meskipun hasil afektif seperti motivasi belajar masih bervariasi antar konteks. Tantangan utama implementasi meliputi keterbatasan literasi korpus guru, infrastruktur teknologi, dan integrasi pembelajaran berbasis korpus dalam kelas EFL. Oleh karena itu, implementasi DDL memerlukan strategi pembelajaran yang kontekstual, fleksibel, dan sesuai dengan kondisi pendidikan lokal.

**Kata Kunci:** *Pembelajaran Berbasis Data, Pedagogi Berbasis Korpus, EFL Pra-Perguruan Tinggi, Kosakata dan Kolokasi, Mediasi Guru*

### ABSTRACT

English as a Foreign Language (EFL) learning in pre-tertiary education is still largely dominated by textbook-based and teacher-centered instruction, limiting learners' exposure to authentic language use. As an alternative, Data-Driven Learning (DDL) offers a corpus-informed approach that enables learners to explore language patterns through authentic language data. This study employed a qualitative analytical literature review approach by conducting a systematic search and selection of articles through Google Scholar and related academic sources based on PRISMA principles. The data were analyzed through thematic synthesis to identify major patterns related to instructional approaches, learning outcomes, and implementation constraints. The findings reveal that DDL is commonly implemented through teacher-mediated and guided discovery approaches that balance learner autonomy with teacher scaffolding. DDL consistently supports vocabulary development, collocational awareness, and language noticing, although affective outcomes such as learner motivation remain inconsistent across contexts. Major implementation challenges include limited teacher corpus literacy, technological constraints, and difficulties integrating corpus-based activities into regular EFL

Copyright (c) 2026 EDUCATOR : Jurnal Inovasi Tenaga Pendidik dan Kependidikan



<https://doi.org/10.51878/educator.v6i2.10856>



instruction. Therefore, effective DDL implementation requires flexible and context-sensitive instructional strategies that are adaptable to local educational conditions.

**Keywords:** *Data-Driven Learning, Corpus-Informed Pedagogy, Pre-Tertiary EFL, Vocabulary and Collocation, Teacher Mediation*

## INTRODUCTION

Data-Driven Learning (DDL) is widely recognized as a corpus-informed pedagogical approach that enables learners to explore authentic language patterns through empirical language data (Johns, 1991; Boulton & Vyatkina, 2021; O’Keeffe & Mark, 2022). Unlike approaches that primarily emphasize rule memorization, DDL encourages learners to identify patterns of form, meaning, and language use through exposure to concordance lines, frequency lists, and collocational evidence derived from language corpora. Within this approach, learners are positioned as active analysts of linguistic evidence, while teachers function as facilitators who guide learners in interpreting authentic language use (O’Keeffe & Mark, 2022). Consequently, DDL has increasingly been associated with inductive learning principles and corpus-informed pedagogies that emphasize discovery learning in second language acquisition contexts (Boulton & Vyatkina, 2021).

Beyond its theoretical foundation, Data-Driven Learning (DDL) has become increasingly relevant in English as a Foreign Language (EFL) education, particularly in contexts where learners have limited access to authentic language input outside the classroom (Crosthwaite et al., 2021). Corpus-informed instruction allows learners to observe how words co-occur, how grammatical patterns vary across contexts, and how frequency influences language use. Empirical studies such as Xue and Liu (2021) demonstrate that DDL is effective in improving lexical awareness, collocational competence, and grammatical development through guided interaction with corpus data. In addition, Yoon and Lee (2024) found that DDL supports learner autonomy by encouraging students to actively investigate language patterns through data-driven inquiry and discovery learning processes. These findings suggest that DDL not only enhances linguistic knowledge but also promotes more independent and reflective language learning practices in EFL contexts.

Despite its pedagogical advantages, research on Data-Driven Learning (DDL) has historically focused more heavily on tertiary education and advanced learners, leaving pre-tertiary EFL contexts such as secondary and vocational education relatively underexplored (Lusta et al., 2023). This imbalance indicates that although DDL is recognized as a corpus-informed approach that supports language awareness and inductive learning, its implementation in school settings remains limited. In many pre-tertiary EFL classrooms, instruction continues to rely heavily on textbooks, explicit grammar explanations, and teacher-centered practices, which restrict learners’ exposure to authentic language data and real patterns of language use in context (Emir & Yangin Ekşi, 2023). As a result, students often have fewer opportunities to engage in exploratory, data-driven learning that could support their lexical and grammatical development. This condition highlights the need for further adaptation of DDL approaches to make them more practical, accessible, and effective in secondary education contexts.

This issue becomes particularly significant in Indonesia, where English is generally learned as a foreign language through formal classroom instruction. In many Indonesian pre-tertiary contexts, access to authentic English input remains limited, while classroom practices continue to emphasize textbook-based and teacher-centered learning (Cahyawijaya et al., 2021). Although corpus-informed approaches such as Data-Driven Learning (DDL) have begun to attract attention in EFL research, their implementation in school settings remains uneven due



to limited technological infrastructure, challenges in teacher training, and difficulties in integrating corpus tools into regular classroom instruction (Larsson et al., 2022). Nevertheless, recent research by Crosthwaite and Meunier (2025) suggests that simplified or teacher-mediated DDL models can provide practical solutions for resource-constrained classrooms, particularly when full corpus software use is not feasible. These findings indicate that although DDL holds strong pedagogical potential, its successful integration into Indonesian EFL classrooms requires careful adaptation to local technological and pedagogical conditions.

This review contributes to the growing body of Data-Driven Learning (DDL) scholarship by specifically synthesizing evidence from pre-tertiary EFL contexts, particularly regarding instructional adaptation, teacher mediation, and context-sensitive implementation challenges. Unlike previous DDL reviews that predominantly focus on tertiary education and general pedagogical effectiveness, this study specifically examines how DDL is adapted and implemented in pre-tertiary EFL settings, especially within resource-constrained educational environments such as Indonesia. Furthermore, unlike many earlier discussions that primarily emphasized university-level contexts, Luo (2025) demonstrates that DDL is increasingly being explored in secondary and non-tertiary educational settings, highlighting both its pedagogical potential and the practical constraints encountered across different learning environments. Accordingly, this review not only highlights reported learning outcomes but also examines contextual factors influencing the sustainability and pedagogical feasibility of DDL in pre-tertiary education. This perspective emphasizes the importance of aligning corpus-informed pedagogy with institutional constraints, teacher readiness, and learner proficiency levels in non-tertiary EFL contexts.

Accordingly, this study aims to synthesize instructional approaches used in pre-tertiary DDL research, examine reported linguistic and affective learning outcomes, and identify persistent implementation challenges in school-level EFL contexts. In addition, the review seeks to map how DDL practices are adapted within classroom environments where technological access, teacher expertise, and curriculum demands may vary considerably. Particular attention is given to studies relevant to the Indonesian context in order to provide context-sensitive insights into how corpus-informed pedagogy can be realistically and sustainably integrated into pre-tertiary English language education. By doing so, this study is expected to contribute not only to the theoretical development of DDL in EFL contexts, but also to provide practical implications for teachers, curriculum developers, and policymakers in improving the feasibility of corpus-based instruction in secondary education settings.

## RESEARCH METHODS

This study employed a qualitative analytical literature review approach to synthesize and critically evaluate previous studies on Data-Driven Learning (DDL) in pre-tertiary English as a Foreign Language (EFL) education. The literature search was conducted systematically through Google Scholar and related academic sources using keywords such as *data-driven learning*, DDL, corpus-based instruction, pre-tertiary EFL, and secondary education. The initial search yielded 30 articles, which were screened through duplicate removal to obtain 28 unique articles. Title and abstract screening resulted in 20 articles relevant to the focus of the study, while irrelevant studies, studies focusing on higher education, and unrelated EFL contexts were excluded. Full-text eligibility assessment was then conducted using inclusion criteria emphasizing explicit discussion of DDL or corpus-based instruction, relevance to pre-tertiary EFL contexts, and theoretical or empirical contributions to language learning research. The selected studies included recent empirical research as well as several seminal conceptual and



methodological studies used as analytical foundations. After one additional article was excluded due to limited access and unclear methodology, a final total of 19 studies were included in the review. The selection process followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) principles to ensure systematic and transparent study identification and selection procedures.

Data analysis was conducted using thematic synthesis to identify recurring patterns across the selected studies. The reviewed articles were systematically categorized into themes related to instructional approaches, reported learning outcomes, and implementation challenges of DDL in pre-tertiary EFL contexts. Each article was examined using a document analysis framework covering research context, educational level, focus of study, research design, and key findings. Relevant data from each study were extracted, compared, and organized to identify similarities and differences across the literature. The synthesized findings were then interpreted narratively to generate comprehensive insights aligned with the objectives of the study. This analytical process enabled the study to provide a clearer understanding of how DDL has been implemented, adapted, and evaluated in different pre-tertiary EFL settings.

## RESULTS AND DISCUSSION

### Results

The literature search was conducted through Google Scholar and related academic sources using keywords such as *data-driven learning*, *DDL*, *corpus-based instruction*, *pre-tertiary EFL*, and *secondary education*. The initial search yielded 30 articles, which were screened through duplicate removal to obtain 28 unique articles. Title and abstract screening resulted in 20 articles relevant to the implementation of Data-Driven Learning (DDL) in pre-tertiary EFL contexts, while 8 articles were excluded due to irrelevance to the research focus. The full-text eligibility stage applied inclusion criteria emphasizing a focus on DDL or corpus-based instruction, relevance to pre-tertiary education, and theoretical or empirical connections to EFL learning. The selected studies included recent empirical research as well as several seminal conceptual and methodological studies used as analytical foundations. After 1 additional article was excluded due to limited access or unclear methodology, a final total of 19 articles were included in the analysis. The selection process followed PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) principles to ensure systematic and transparent identification and selection procedures. A summary of the reviewed studies is presented in Table 1.

**Table 1. Summary of Reviewed Studies on Data-Driven Learning (DDL) in Pre-Tertiary EFL Education**

Autho & Year	Context / Country	Education al Level	Focus of Study	Research Design	Key Reported Outcomes
Johns (1991)	Global	Conceptual	Foundational concept of DDL	Conceptual	Introduced learner-as-researcher principle
McEneaney & Hardie (2012)	Global	General education	Corpus linguistics foundations	Conceptual	Corpora as empirical evidence for

					language analysis
Ellis (2002)	Global	SLA	Noticing & form-meaning mapping	Theoretical	Repeated patterned input supports acquisition
Meunier (2020)	Europe	General education	Scaffolding & sustainability of DDL	Conceptual	Emphasizes constructive alignment in DDL
Boulton (2020)	Global	Secondary-tertiary	DDL pedagogy & accessibility	Review	DDL feasible beyond advanced learners
O’Keeffe (2021)	Global	General education	Broadening DDL research gaze	Conceptual review	Calls for expansion to school-level contexts
Boulton & Vyatkina (2021)	Global	Pre-tertiary-tertiary	DDL principles & pedagogy	Review	Highlights guided discovery & scaffolding
Crosthwaite, Luciana, & Schweinberger (2021)	Indonesia	Pre-service teachers	Teacher perceptions of DDL	Mixed-methods	Mixed readiness; secondary more receptive
Asi et al. (2021)	Indonesia	EFL learners (non-university)	Sustainability of DDL practices	Longitudinal survey	Continued autonomous DDL use
Bowen (2009)	Global	Research methods	Document analysis	Methodological	Validates document analysis as research tool
Braun & Clarke (2006)	Global	Research methods	Thematic analysis	Methodological	Framework for qualitative synthesis

Grant & Booth (2009)	Global	Research methods	Review typology	Methodological review	Defines analytical literature reviews
Snyder (2019)	Global	Research methods	Literature review methods	Methodological	Guides thematic synthesis in reviews
Rifaldo, Maretha, & Pradita (2022)	Indonesia	Senior high school	DDL for descriptive text	Qualitative classroom study	Teacher-mediated DDL feasible
Phoocharoensil (2022)	Thailand	Secondary-tertiary	Teacher perceptions of DDL	Qualitative	DDL beneficial but ICT-demanding
Zare & Karimpor (2022)	Iran	Secondary education	DDL & motivation	Quasi-experimental	No significant motivational difference
He & Xie (2024)	China	Pre-tertiary	DDL for collocation learning	Experimental	Sustained collocation gains
Saputro & Hima (2024)	Indonesia	Vocational high school	DDL & engagement	Quasi-experimental	Improved learning engagement
Huang & Ma (2025)	Global	Pre-tertiary	DDL effectiveness & perceptions	Systematic review	Effective across ages with scaffolding

Table 1 presents a summary of the reviewed studies on Data-Driven Learning (DDL) in pre-tertiary EFL education across various educational contexts and research designs published between 1991 and 2025. The reviewed literature includes conceptual, theoretical, methodological, review-based, and empirical studies that collectively provide the theoretical and analytical foundation for understanding DDL implementation in EFL settings. Foundational studies emphasize learner-centered language exploration, corpus-based analysis, and the role of noticing in language acquisition, while recent empirical studies increasingly focus on secondary and pre-tertiary contexts in countries such as Indonesia, Thailand, Iran, and China. The studies mainly investigate instructional practices, teacher perceptions, learner engagement, vocabulary and collocation learning, as well as the sustainability of DDL implementation. Overall, the findings consistently indicate that DDL supports vocabulary development, collocational awareness, learner autonomy, and guided discovery learning, although challenges related to technological limitations, ICT demands, and learner readiness remain significant in many educational contexts.



### **DDL Approaches and Interaction Patterns in Pre-Tertiary EFL Contexts**

The reviewed studies demonstrate that Data-Driven Learning (DDL) in pre-tertiary EFL education is generally implemented through two dominant approaches, namely teacher-mediated DDL and learner-operated DDL. Among the analyzed studies, teacher-mediated approaches appeared more frequently, particularly in secondary schools and vocational education settings. In these implementations, teachers commonly selected and simplified corpus materials before presenting them to learners through worksheets, concordance lines, frequency lists, and classroom presentations. This approach was mainly applied in contexts where students had limited experience with corpus tools or where schools faced technological limitations such as restricted internet access and insufficient digital devices. The findings also indicate that teacher mediation was frequently used to reduce cognitive load and help learners focus on targeted linguistic features during classroom activities.

In contrast, learner-operated DDL approaches were reported less frequently across the reviewed studies. In these approaches, students interacted directly with corpus tools to investigate language patterns and complete guided discovery tasks. Most implementations of learner-operated DDL involved structured instruction and teacher guidance to support learners in interpreting concordance outputs and corpus evidence. Collaborative learning activities such as pair work and small-group discussion were also commonly integrated to facilitate peer interaction and reduce learner difficulty during corpus exploration. In several classroom contexts, whole-class interaction using projected corpus data was used as an alternative strategy in situations where individual device access was unavailable. These findings indicate that DDL implementation in pre-tertiary contexts is highly adaptive and influenced by learner readiness, classroom conditions, and technological availability.

### **Reported Learning Outcomes of DDL for Pre-Tertiary Learners**

The reviewed studies consistently reported vocabulary development and collocation awareness as the most dominant learning outcomes associated with DDL implementation. Learners who participated in corpus-based activities generally demonstrated improved understanding of vocabulary usage patterns and contextual language use. Corpus exploration activities allowed learners to observe repeated lexical patterns across authentic language examples, particularly in grammar and writing-related tasks. Several studies also reported improvements in learners' ability to recognize appropriate collocations and contextual word combinations after participating in DDL-based instruction. These outcomes were identified across different educational levels, including secondary schools and vocational education contexts.

In addition to linguistic outcomes, some studies reported improvements related to learner autonomy and independent learning behavior. Learners gradually developed strategies for investigating language problems independently through corpus consultation activities and guided discovery tasks. A number of studies also found that learners continued using corpus-based strategies outside formal classroom instruction, indicating the potential for sustainable learning habits. However, affective outcomes such as motivation and classroom engagement showed less consistent patterns across the reviewed literature. While some studies described positive learner attitudes toward DDL activities, others reported that motivational outcomes depended heavily on task design, learner proficiency, and classroom relevance. Overall, cognitive and language-related outcomes were reported more consistently than affective outcomes. To provide a clearer synthesis of the major findings across the reviewed studies, Table 2 summarizes the dominant themes, learning outcomes, and implementation challenges identified in pre-tertiary DDL research.

**Table 2. Major Themes and Patterns of DDL Implementation in Pre-Tertiary EFL Education**

Theme	Main Findings	Supporting Studies
Instructional Approaches	DDL was predominantly implemented through teacher-mediated and guided discovery approaches	Meunier (2020); Boulton & Vyatkina (2021); Rifaldo et al. (2022)
Learning Outcomes	DDL improved vocabulary development, collocational awareness, and language noticing	He & Xie (2024); Saputro & Hima (2024); Huang & Ma (2025)
Learner Autonomy	Learners gradually developed independent language exploration strategies through corpus consultation	Asi et al. (2021); Yoon & Lee (2024)
Affective Outcomes	Motivation and engagement outcomes remained inconsistent across contexts	Zare & Karimpour (2022); Saputro & Hima (2024)
Teacher Challenges	Teachers experienced difficulties related to corpus literacy and task design	Crosthwaite et al. (2021); Phoocharoensil (2022)
Technological Constraints	Limited internet access and insufficient digital devices hindered DDL implementation	Rifaldo et al. (2022); Phoocharoensil (2022)
Sustainability Issues	Effective long-term implementation required scaffolding and contextual adaptation	Meunier (2020); Huang & Ma (2025)

Table 2 demonstrates that the reviewed studies converge around several recurring themes related to instructional practices, learning outcomes, and implementation challenges in pre-tertiary DDL contexts. Teacher-mediated and guided discovery approaches appeared to dominate classroom implementation due to learner readiness and technological limitations in many school environments. Across studies, vocabulary development, collocational awareness, and language noticing emerged as the most consistently reported linguistic outcomes, while affective outcomes such as learner motivation showed more context-dependent results. The table also highlights that successful DDL implementation depends not only on technological access, but also on teacher corpus literacy, pedagogical scaffolding, and the ability to adapt corpus-informed activities to local classroom conditions.

### Challenges and Constraints in DDL Implementation

Several recurring challenges related to DDL implementation were identified across the reviewed studies. One of the most frequently reported issues concerned teacher corpus literacy, particularly teachers' limited familiarity with corpus tools and uncertainty regarding effective classroom implementation strategies. Both pre-service and in-service teachers often recognized the potential value of DDL but expressed concerns regarding learner readiness, classroom management, and pedagogical feasibility. These concerns appeared more prominent in lower educational levels where students were considered less familiar with autonomous learning practices and analytical language tasks.



Technological limitations also emerged as a significant challenge in many pre-tertiary educational contexts. Limited access to computers, digital devices, and stable internet connections restricted opportunities for learner-operated DDL activities in many schools. As a result, teachers frequently relied on low-technology or teacher-mediated forms of corpus instruction. Another important issue involved task design and cognitive load. Several studies reported that learners experienced confusion when corpus tasks presented excessive amounts of data without sufficient instructional guidance or scaffolding. Furthermore, the review revealed that most existing studies focused primarily on short-term classroom interventions, while relatively few studies investigated the long-term effects of DDL on language development and learner autonomy.

## **Discussion**

### **DDL as a Guided Discovery Approach in Pre-Tertiary EFL Education**

The findings of this review indicate that Data-Driven Learning (DDL) in pre-tertiary EFL contexts is more effectively implemented through teacher-mediated and guided discovery approaches rather than through fully autonomous corpus consultation. This pattern appears consistently across the studies summarized in Table 1 and Table 2, where classroom implementation was generally supported through teacher scaffolding, simplified concordance materials, worksheets, and collaborative learning activities. These findings align with constructivist learning principles, which emphasize that learners develop understanding more effectively when instructional support and structured assistance are provided during the learning process. The dominance of teacher-mediated approaches in secondary and vocational school contexts further suggests that pre-tertiary learners still require substantial pedagogical support to interpret corpus evidence meaningfully and participate effectively in corpus-based learning activities.

This interpretation is consistent with previous studies arguing that DDL should not be viewed as a rigid technological method, but rather as a flexible pedagogical orientation that can be adapted according to learner readiness and classroom conditions (Boulton & Vyatkina, 2021). The reviewed studies also demonstrate that successful DDL implementation is influenced more strongly by pedagogical mediation, scaffolding quality, and task organization than by technological sophistication alone. Lin (2021) emphasized that the effectiveness of DDL largely depends on how teachers design learning tasks and facilitate learner interaction with corpus materials. Similarly, Tosun and Sofu (2023) highlighted the importance of guided concordancing activities in helping EFL learners develop language awareness and interpret corpus evidence more effectively. Therefore, these findings reinforce the view that DDL implementation in pre-tertiary contexts should be introduced gradually through contextualized and structured instructional models.

The findings further reveal that DDL retains pedagogical value even in educational settings with limited technological infrastructure. Several reviewed studies reported that teachers adapted DDL instruction through printed worksheets, projected corpus examples, pair work, and whole-class corpus exploration when individual device access was unavailable. These adaptations demonstrate that instructional flexibility is a crucial factor in sustaining DDL implementation, particularly in Indonesian school contexts. This interpretation aligns with Pérez-Paredes (2022), who argued that the pedagogical value of corpus-based learning is determined more by instructional design than by technological complexity. Likewise, Liu and Ma (2025) emphasized that low-technology corpus integration can still support meaningful language discovery when teachers provide structured guidance and context-sensitive learning



activities. Consequently, the feasibility of DDL implementation in pre-tertiary education should primarily be understood as a pedagogical and contextual issue rather than merely a technological one.

### **Interpretation of Learning Outcomes**

The findings of this review consistently show that vocabulary development, collocational awareness, and language noticing are the most dominant learning outcomes associated with DDL implementation in pre-tertiary EFL contexts. Across the reviewed studies, learners participating in corpus-based activities demonstrated improved understanding of contextual vocabulary use, collocational patterns, and grammatical forms through repeated interaction with authentic language examples. These findings were particularly evident in studies focusing on collocation learning, descriptive writing, and vocabulary development, where learners benefited from observing authentic lexical patterns through concordance lines and corpus data. The results reinforce the theoretical assumption that corpus-based learning provides repeated and contextualized exposure to language patterns, enabling learners to recognize linguistic regularities more effectively than through isolated rule explanation.

This interpretation supports previous studies suggesting that authentic corpus evidence enables learners to identify semantic relationships, collocations, and language-use patterns that are often difficult to acquire through conventional textbook-based instruction alone (Satake, 2022). In this sense, DDL contributes not only to language acquisition but also to learners' awareness of how language functions in authentic communicative contexts. These findings further strengthen the view that corpus-informed pedagogy can complement traditional language teaching approaches by providing more contextualized and meaningful exposure to language use. Moreover, the emphasis on guided discovery learning across the reviewed studies indicates that DDL supports inductive learning processes, in which learners actively construct linguistic understanding through exploration and pattern identification.

In addition to linguistic outcomes, several reviewed studies also reported improvements related to learner autonomy and independent learning behavior. Learners gradually developed strategies for investigating language-related problems independently through corpus consultation activities, concordance analysis, and guided inquiry tasks. Some studies further indicated that learners continued using corpus-based strategies outside formal classroom instruction, suggesting the potential development of sustainable language learning habits. These findings are consistent with Kalyaniwala and Ciekanski (2021) and Ludwig and Tassinari (2021), who argued that repeated engagement with corpus evidence can foster analytical learning habits and independent language-learning strategies.

However, the review also demonstrates that affective outcomes such as learner motivation and classroom engagement remain inconsistent across research contexts. Some studies reported positive learner attitudes toward DDL activities and increased classroom participation, whereas others found no significant motivational differences compared to conventional instruction. These inconsistencies suggest that affective outcomes are strongly influenced by instructional design, learner readiness, task relevance, and language proficiency rather than by DDL implementation itself. Therefore, motivation and engagement should not be generalized as automatic outcomes of corpus-based learning. Instead, the effectiveness of DDL in generating positive affective responses depends heavily on how corpus activities are aligned with learners' cognitive readiness and instructional needs.



### **Implications for English Teacher Education in Indonesia**

The findings of this review highlight the importance of strengthening corpus literacy and pedagogical competence within English teacher education programs in Indonesia. Since many Indonesian classrooms remain dominated by textbook-based and teacher-centered instruction, the integration of DDL requires teachers to develop both technological understanding and corpus-informed pedagogical strategies. The reviewed studies indicate that teachers generally recognize the pedagogical potential of DDL but still experience uncertainty regarding classroom implementation, task design, learner readiness, and corpus tool usage. These findings suggest that teacher education programs should not merely introduce corpus tools at a technical level, but should also provide practical training on integrating corpus-informed activities into authentic classroom situations.

Experiential learning opportunities are particularly important because they allow pre-service teachers to experience DDL from the learner perspective before implementing it in professional teaching practice. Oktavianti et al. (2023) also emphasized the importance of providing authentic experiences with corpus-informed instructional materials for future teachers. The findings further indicate that Indonesian schools may be more suited to simplified and context-sensitive DDL approaches rather than highly technology-dependent implementation models. In many reviewed contexts, teacher-mediated activities such as worksheet-based concordancing, projected corpus data, and collaborative analysis represented more realistic forms of implementation for classrooms with limited resources.

These findings reinforce previous research suggesting that successful educational innovation depends largely on contextual adaptation rather than direct replication of technologically advanced models from different educational environments. Therefore, teacher professional development programs should place greater emphasis on pedagogical integration, scaffolding strategies, and classroom task design rather than focusing solely on technical corpus training. Sari et al. (2021) similarly argued that teacher professional development should support the integration of technological, pedagogical, and content knowledge within authentic classroom contexts. Consequently, sustainable DDL implementation in Indonesia requires not only technological access, but also long-term institutional support, context-sensitive teacher preparation, and adaptive instructional models suited to local classroom realities.

### **Contribution, Limitations, and Future Research Directions**

This review contributes to the growing discussion of DDL implementation in pre-tertiary EFL education by synthesizing recent evidence related to instructional adaptation, learning outcomes, and implementation challenges across different educational contexts. Unlike many previous discussions that primarily emphasized the effectiveness of DDL in improving language skills, this review demonstrates that successful DDL implementation is strongly associated with teacher mediation, scaffolding quality, learner readiness, and contextual feasibility. The findings also provide a clearer understanding of how DDL practices are adapted within resource-constrained educational environments, particularly in developing countries such as Indonesia. In addition, this review more clearly distinguishes between cognitive outcomes consistently supported across studies and affective outcomes that remain contextual and inconclusive. This distinction is important in preventing overly generalized assumptions regarding the benefits of DDL implementation (Drajati et al., 2021).

Nevertheless, several limitations of this review should be acknowledged. First, most of the reviewed studies were dominated by conceptual discussions, review articles, and small-scale classroom interventions, meaning that empirical evidence regarding DDL implementation in pre-tertiary EFL contexts remains relatively uneven. Second, DDL implementation across



the reviewed studies varied considerably in terms of instructional approaches, levels of teacher involvement, and technological integration, making direct comparison across studies difficult. Third, affective outcomes such as learner motivation and engagement also demonstrated inconsistent findings because they were highly influenced by task design, learner readiness, and classroom conditions. Furthermore, most existing studies primarily focused on vocabulary and collocation development, while research investigating the impact of DDL on other language skills and its long-term sustainability in classroom practice remains limited.

Therefore, future research should prioritize longitudinal and classroom-based studies examining the sustained impact of DDL on vocabulary growth, collocational competence, learner autonomy, and pedagogical adaptation over longer periods of time. Future studies should also explore how different scaffolding models, corpus-task designs, and technology integration strategies influence learner engagement and instructional effectiveness across diverse educational contexts. In addition, studies specifically focusing on Indonesian secondary schools and vocational education settings would provide more context-sensitive insights into how DDL can be sustainably integrated into pre-tertiary EFL instruction. Such findings are expected to contribute to the development of more practical, adaptive, and contextually appropriate corpus-informed pedagogies for diverse classroom realities.

## **CONCLUSION**

Overall, it can be concluded that Data-Driven Learning (DDL) has significant potential to support English as a Foreign Language (EFL) learning in pre-tertiary education. This review demonstrates that DDL is most effectively implemented through teacher-mediated and guided discovery approaches, in which teachers play an essential role in providing scaffolding and guiding learners in exploring authentic language data. The findings also indicate that DDL consistently supports vocabulary development, collocational awareness, and language noticing, while simultaneously encouraging learner autonomy through corpus-based language exploration activities.

On the other hand, the implementation of DDL still faces several challenges, particularly related to teachers' corpus literacy, technological infrastructure, task design, and learner readiness. Therefore, DDL should not be viewed as a standalone method, but rather as a flexible pedagogical approach that can be adapted to different learning contexts. In the Indonesian context, the successful integration of DDL requires stronger teacher pedagogical competence, context-sensitive instructional strategies, and realistic corpus integration aligned with school conditions. Thus, this study confirms that the effectiveness of DDL largely depends on teacher mediation, instructional scaffolding, and the ability to adapt corpus-informed pedagogy to the needs of pre-tertiary EFL classrooms.

## **REFERENCES**

- Boulton, A., & Vyatkina, N. (2021). Thirty Years Of Data-Driven Learning: Taking Stock And Charting New Directions Over Time. *Language Learning & Technology*, 25(3), 66–89. <https://www.lltjournal.org/item/10125-73450/>
- O’Keeffe, A., & Mark, G. (2022). Principled Pattern Curation To Guide Data-Driven Learning Design. *Applied Corpus Linguistics*, 2(3), 100028. <https://doi.org/10.1016/j.acorp.2022.100028>
- Cahyawijaya, S., Winata, G. I., Wilie, B., Vincentio, K., Li, X., Kuncoro, A., ... & Fung, P. (2021). IndoNLG: Benchmark and resources for evaluating Indonesian natural language



generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 8875-8898).

<https://doi.org/10.48550/arXiv.2104.08200>

Crosthwaite, P., & Meunier, F. (2025). When DDL Goes Wrong: The Importance Of Constructive Alignment In Secondary Education. *The EuroCALL Review*, 32(1), 45–63.

<https://doi.org/10.4995/eurocall.2025.23844>

Crosthwaite, P., Wijaya, D., & Luciana. (2021). 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>

Drajati, N. A., Tan, L., Haryati, S., Rochsantiningsih, D., & Zainnuri, H. (2021). Investigating English language teachers in developing TPACK and multimodal literacy. *Indonesian Journal of Applied Linguistics*, 10(3), 575–585. <https://doi.org/10.17509/ijal.v7i3.9806>

Emir, G., & Yangın Ekşi, G. (2023). Corpus Used As A Data-Driven Learning Tool In L2 Academic Writing: Evidence From Turkish Contexts. *TEFLIN Journal*, 34(2), 209–225.

<https://doi.org/10.15639/teflinjournal.v34i2/209-225>

Kalyaniwala, C., & Ciekanski, M. (2021). Autonomy Calling: A Systematic Review Of 22 Years Of Publications In Learner Autonomy And CALL. *Language Learning & Technology*, 25(3), 106–131. <https://www.lltjournal.org/item/10125-73452/>

Larsson, T., Staples, S., & Egbert, J. (2022). Teaching, Learning, And Researching With Corpora. *Applied Corpus Linguistics*, 2(3), 100025.

<https://doi.org/10.1016/j.acorp.2022.100025>

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>

Liu, J., & Ma, Q. (2025). Examining Corpus-Based Language Pedagogy (CBLP) Practices In Data-Driven Learning (DDL) For Low-Proficiency L2 English Learners: A Meta-Analysis. *Educational Technology & Society*, 28(2), 53–76.

[https://doi.org/10.30191/ETS.202504\\_28\(2\).RP04](https://doi.org/10.30191/ETS.202504_28(2).RP04)

Ludwig, C., & Tassinari, M. G. (2021). Foreign Language Learner Autonomy In Online Learning Environments: The Teachers' Perspectives. *Innovation in Language Learning and Teaching*, 17(2), 217–234. <https://doi.org/10.1080/17501229.2021.2012476>

Luo, H. (2025). Data-Driven Learning In Second Language Writing: A Systematic Review Of Efficacy, Challenges, And Future Directions. *International Journal of Education and Humanities*. <https://doi.org/10.54097/t1wp5032>

Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language Corpus And Data-Driven Learning (DDL) In Language Classrooms: A Systematic Review. *Heliyon*, 9(12), e22731. <https://doi.org/10.1016/j.heliyon.2023.e22731>

Oktavianti, I. N., Eriani, E., Rolyana, I., & Prayogi, I. (2023). Investigating The Use Of Corpus-Informed Grammar Materials In Indonesian EFL Classrooms. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 7(2), 417–438.



<https://doi.org/10.21093/ijeltal.v7i2.1415>

- Pérez-Paredes, P. (2022). A Systematic Review Of The Uses And Spread Of Corpora And Data-Driven Learning In CALL Research During 2011–2015. *Computer Assisted Language Learning*, 35(1–2), 36–61. <https://doi.org/10.1080/09588221.2019.1667832>
- Sari, Y. R., Drajati, N. A., So, H.-J., & Sumardi, S. (2021). Enhancing EFL Teachers' Technological Pedagogical Content Knowledge (TPACK) Competence Through Reflective Practice. *TEFLIN Journal*, 32(1), 117–133. <https://doi.org/10.15639/teflinjournal.v32i1/117-133>
- Satake, Y. (2022). The Effects Of Corpus Use On L2 Collocation Learning. *The JALT CALL Journal*, 18(1), 34–53. <https://doi.org/10.29140/jaltcall.v18n1.520>
- Tosun, S., & Sofu, H. (2023). The Effectiveness Of Data-Driven Vocabulary Learning: Hands-On Concordancing Through A Pedagogical Corpus. *Journal of Language and Education*, 9(3), 176–190. <https://doi.org/10.17323/jle.2023.12426>
- Xue, L., & Liu, F. (2021). Using Data-Driven Learning Activities To Improve Lexical Awareness In Intermediate EFL Learners. *Cogent Education*, 8(1), 1996867. <https://doi.org/10.1080/2331186X.2021.1996867>
- Yoon, K.-H., & Lee, D. J. (2024). A Meta-Analysis Of Data-Driven Learning (DDL) In EFL/ESL Settings. *Brain, Digital, & Learning*, 14(2), 283–304. <https://doi.org/10.31216/BDL.20240017>