

JOURNAL OF RESEARCH STUDIES IN ENGLISH LANGUAGE TEACHING AND LEARNING



To cite this article in APA 7th style:

Worasit, T., & Boonchai, C. (2025). Effectiveness of data-driven learning (DDL) approaches. *Research Studies in English Language Teaching and Learning*, 3(1), 1–17. <https://doi.org/10.62583/rseltl.v3i1.70>

More Citation Formats

For more citation styles, please visit: <https://rseltl.pierreonline.uk/>

This article is published by **Pierre Online Publications Ltd**, a UK publishing house

Effectiveness of data-driven learning (DDL) approaches

Thongprasert Worasit¹ & Chanan Boonchai²

¹Udon Thani Rajabhat University, Thailand

Abstract

This study examines the effectiveness of Data-Driven Learning (DDL) in enhancing grammar instruction for Thai English as a Foreign Language (EFL) learners, addressing the limitations of traditional methods such as the Grammar Translation Method (GTM). DDL, rooted in corpus linguistics, engages learners in analysing authentic language patterns through concordancing tools and corpus-based exercises, promoting grammatical accuracy, lexical proficiency, and learner autonomy. Using a quasi-experimental design, 60 undergraduate students were divided into an experimental group, taught using DDL, and a control group, adhering to conventional methods. The six-week intervention included pre- and post-tests to measure language proficiency and Likert-scale questionnaires to assess learner engagement. Results indicated significant improvements in the experimental group's grammatical competence and satisfaction with the learning process, highlighting DDL's capacity to bridge the gap between knowledge and application. Despite challenges such as low-proficiency learners navigating authentic data and increased teacher preparation, scaffolding and resource development were effective solutions. The findings underscore DDL's potential to transform EFL classrooms, particularly in contexts with structural language differences and systemic challenges, fostering active and meaningful learning. This study contributes to the growing body of research advocating innovative, learner-centred approaches, offering valuable insights for educators and policymakers seeking to modernise grammar instruction.



ISSN (online): 2977-0394

KEYWORDS

data-driven learning, grammar instruction, corpus linguistics, EFL learners, Thai education



Under Creative Commons Licence:
Atribución 4.0 Internacional (CC BY 4.0)

Introduction

Grammar teaching has long been recognised in EFL contexts to be one of the most complicated and complex areas of language teaching. Grammar is a basic communicational resource for any speaker, yet it is one domain where many learners, especially then moving from rule-based knowledge to practical application, face difficulties. More traditionally, methods such as GTM have dominated EFL classroom practice. Such methods, though useful for the effective mastering of grammatical rules, are ineffective in real-life communication for the learners. The gulf that such a gap created between grammatical proficiency and communicative ability demands the designing of innovative approaches in order to bring such a disconnect to a meaningful close.

Recent decades have indeed provided a number of new possibilities introduced by corpus linguistics for teaching and learning. It is the very concept denoting analysis, through the computer, of naturally occurring language data stored in electronic corpora, which now underlies actual language usage, its frequency, and contextual meaning—a rather serious revision from purely prescriptive grammatical rules for detailed views on how an utterance might function in many contexts. DDL is an instruction-driven pedagogical approach based on corpus linguistics. First theorised by Tim Johns in the 1980s, DDL integrates corpus methodologies right within the language classroom, where it is the learners themselves who are led towards the induction of linguistic generalisation by exploring original language data. The appeal of DDL lies in its compatibility with other related current SLA theories such as Krashen's comprehensible input hypothesis and Schmidt's noticing hypothesis. Each of these different theories underlines the meaningfulness of exposure to language and engagement in the process of learning. By allowing learners to interact directly with authentic language data, DDL encourages them to notice and internalise grammatical and lexical structures in context. It not only helps in enhancing grammatical accuracy but also in critical thinking, learner autonomy, and long-term retention.

Despite its theoretical strengths, the actual use of DDL in the EFL classroom has been relatively scanty, especially within educational settings which still rely more on traditional methods of teaching. In Thailand, for example, grammar instruction is frequently typified by the use of deductive methods wherein learners are presented with grammatical rules before they practice them through controlled exercises. Though it is indeed a way for them to score well in tests and examinations, the method does not significantly improve the proficiency of learners to use grammar accurately and meaningfully in real situations. Besides that, structural features present in Thai cause additional complications. The Thai language has no inflection of verb tense, for example, or articles—grammatical components peculiar and essential to English. These conceptual differences can get very troublesome, leading to persistent errors and a lack of confidence in grammatical usage.

Add to these systemic issues in Thai EFL education, there are a number of systemic issues related to class sizes, textbooks, and general lack of exposure to the use of the language in naturalistic situations. This usually culminates into passive learning whereby learners do not get ample opportunities to participate or engage in critical thinking. What is needed for a change in these issues is interactive engagement of students in learning with more learner-centred approaches. That is where DDL opens a very promising perspective: by using corpus linguistics insights and including authentic language data into the classroom, grammar could be taught in a far more effective and captivating way.

This study attempts to investigate the effectiveness of DDL in developing grammatical competence and lexical knowledge among Thai EFL learners. In a quasi-experimental design, this study compares the results of an experimental group that received instruction through DDL approaches with a control group receiving traditional teaching methods. The intervention lasted for six weeks, with both groups participating in instructional sessions designed to align with the programme's learning outcomes. The experimental group receives corpus-based materials and concordancing tools to look for the revelation of authentic language patterns in data-driven exercises, whereas the control group is given a traditionally driven curriculum that places great emphasis on memorisation of rules and rote learning. Some of the key objectives guiding this study include ascertaining whether DDL can significantly enhance grammatical correctness and lexical ability among Thai EFL learners more than traditional methods. It also seeks to explore the impact that DDL might have on learner engagement and satisfaction, particularly in addressing a systemic challenge of large class sizes and unengaging content. Finally, the study investigates the pragmatism of classroom implementation of DDL in Thai EFL classrooms, including the preparation of teachers and materials and the accessibility to learners.

By addressing these objectives, the research contributes to an increasing volume of studies dealing with the pedagogical applications of corpus linguistics and DDL. Although previous studies have established the effectiveness of DDL in various contexts, more context-specific research is called for, given the particular challenges and opportunities accompanying different educational settings. The implication of this study goes beyond the test of the effectiveness of DDL to give a realistic view of how to apply DDL in a Thai EFL context. The study has wider implications for language teaching, especially for regions which are still predominantly relying on traditional approaches and that have limited resources to invest in newer methodologies.

The outcome of this study is important to inform best practices in grammar teaching in EFL settings. DDL embodies a new direction in language instruction through the interrelation between corpus-based methodologies and learner-centred teaching strategies. It oversteps the limitations imposed by traditional approaches and offers a holistic, dynamic framework for developing grammatical competence. With the continuous change that language

education faces due to globalisation and developments in technology, methodologies such as DDL stand out in the development of the future of teaching and learning.

This research, therefore, fills the significant gap in EFL grammar teaching by exploring the effectiveness of DDL in a Thai educational context. Based on both theoretical insights and practical applications, it is expected to provide an overall assessment of the effectiveness of DDL as a novel approach to language teaching. The results could have a general impact on the study's findings to affect Thai EFL education and further enrich the discipline of language pedagogy, besides serving as an important lesson for teachers, researchers, and policy makers in general.

Literature review

Effectiveness of data-driven learning (DDL) approaches

Corpus linguistics and data-driven learning

Corpus linguistics is the study and analysis of naturally occurring language data stored in electronic corpora. Corpus linguistics allows for both qualitative and quantitative research of patterns of language, including frequency counts and contextual analyses. Corpus linguistics, according to Lindquist (2009), is a methodology to analyse linguistic rules and changes over time which may provide insights into authentic use of language. Based on this, Data-Driven Learning (DDL), started by Tim Johns in the 1980s, uses corpus methodologies directly in teaching. As suggested by its name, DDL proposes that learners explore authentic data on their own to come to an inductive discovery of linguistic patterns (Johns, 1991).

DDL is also compatible with SLA theories such as Krashen's comprehensible input hypothesis and Schmidt's noticing hypothesis because the learners are led to actively discover the grammatical and lexical structures themselves. It also allows grammar and lexis to be integrated more smoothly, and learners develop a better understanding of how language is used contextually.

DDL in language teaching

Indirect and Direct Applications of DDL Indirect and direct uses of DDL in teaching have been discussed. Römer (2008), for example, mentions its possible use in designing corpus-based syllabi and materials that may highlight the discrepancies in the content between textbooks and genuine use. In fact, textbooks are often without illustrative examples typical of authentic nature and do not give prominence to grammatical patterns of high frequency, which can be supplied with corpora assistance. For instance, corpus-informed textbooks, such as the "Touchstone" series, incorporate authentic language data that allows learners to become more involved and understand the material better.

On the other hand, direct applications of DDL include independent learners examining linguistic patterns through concordance lines and corpus tools. He was of the view that this would keep active learning and long-term retention at work because the learner would have become a "language detective" (Johns, 1997, as cited in Gilquin

& Granger, 2010). Such methods also correspond to the task-based and communicative approaches to the teaching of languages, which emphasise collaboration and autonomy (Hughes, 2010).

Corpus as a tool for language investigation

Corpora offer special advantages for language learning, especially in their ability to provide authentic data for inductive analysis. According to Lindquist (2009), this is attributed to speed and reliability during corpus analysis in uncovering fine details of patterns of use that might be missed by native speaker intuition. Concordance lines and frequency lists help learners identify recurring patterns and co-occurrence based on some more tangible evidence, which in turn provides them with a concrete basis for the discovery of language (Evison, 2010; Sripicharn, 2002). Jones and Waller (2015) say that corpora challenge traditional rules since real usage is shown, which often goes against prescriptive norms, and hence leads to a more realistic understanding of language.

According to Johns (1991), the direct types of DDL imply that a learner, on their own initiative, discovers linguistic patterns with the help of concordance lines and corpus tools. The data-driven nature of such an approach, therefore, encourages active learning and long-term retention by making learners "language detectives" (Johns, 1997, as cited in Gilquin & Granger, 2010). Another important advantage is its consistency with task-based and communicative approaches to language teaching, including collaborative and autonomous dimensions (Hughes, 2010).

Evidence from previous studies

Empirical investigations underline the efficiency of DDL in language learning. The study by Boulton in 2010 investigated whether paper-based DDL outperforms conventionally addressed alternatives in the achievement of grammatical competence and in changing some of the wrong assumptions among 30 lower-intermediate learners over a four-day period. Conversely, Yunus and Awab, (2012), established that in using DDL on prepositional collocations learning outcome, Malaysian students were found out to be higher, as compared with when adopting conservative approaches. Overall, it does suggest a greater depth and quality in terms of how, through DDL exposure, the authenticity of genuine patterns is communicated to learners, thus giving learning a broader support area. Smart (2014) further highlighted DDL's effectiveness in promoting guided inductive learning, which helped to retain information for a more extended period. Also, Huang (2014) referred to the efficiency of the approach in developing lexico-grammatical patterns in the L2 writing. Both studies underline again the importance DDL has for encouraging participation and contextualising, which is relevant for efficient language learning. Melding corpus linguistics with an advanced pedagogy, DDL holds great hope of responding to most of the limiting factors

in tradition-derived language teaching at schools and reinforces grammatical and lexical competence on an authentic and functional basis (Chingchit, 2008; Choomthong, 2014).

Problems of teaching grammar to Thai learners

Grammar remains one of the most difficult English language items to teach to Thais, partly because of the fundamental structural differences between Thai and English. According to Anyan (2006), Thai grammar is less complicated than that of English, given that it does not have verb tense that changes to afflict the time and the use of articles which Thai does not have. These result in important barriers for Thai learners, who often fail to choose the correct English grammar or construct grammatically correct sentences of their own. Many Thai students who have been taught grammar from early childhood still cannot appropriately use their knowledge in effective communication. But now, Thai teachers are further compounding the entire process by being rather persistent in their methodology of deductive methods, in particular, of the Grammar Translation Method. Despite the fact that in any circumstances, GTM may prove rather promising due to possible positive results only in certain contextual backgrounds communicative competence under such a type of method won't be brought into play since it entirely focuses on keeping in mind gram lemmas' rules and then translating them in context. Because of this type of learning factor, an ideal balance needs to be developed whereby an integration of a deductive versus inductive manner combines alongside placing an extra emphasis on an approach that caters to Student-centred learning principles.

Data-driven learning (DDL) as a solution

Inductive teaching methods, such as Data-Driven Learning, on the other hand, allow students to discover grammar themselves with the help of authentic language data. DDL is based on the use of concordance lines drawn from corpora that show native speakers using the language in context. Learners analyse and generalise patterns of language, developing autonomy and critical thinking. The teacher thus acts as a facilitator, guiding students in their language investigations. This student-initiated research model aims to empower learners and build case-hardened confidence in grammar usage.

Classroom issues and how to overcome them using DDL

Besides that, Thai classes usually have additional obstacles to overcome: huge class sizes, sometimes with as many as 60–80 students in a single section, coupled with outdated textbooks and uninteresting content. This is discouraging for students' participation and even relating the lessons learned to more relevant contexts. Indeed, many EFL learners in Thailand hardly get the chance to practice the language authentically; therefore, it contributes to worsening all of these conditions. Reliance on GTM and deductive approaches on the teacher's part

reduces possibilities of interactive learning. Direct questioning methods may further make students more fearful and squeeze their confidence to participate in discussions.

This also helps to facilitate overcoming difficulties in grammar classes in areas such as less individual confidence due to large class sizes-a problem partly helped by DDL's promotion of small-group work with a paper-based basis for materials created from corpora to enhance awareness of ways that genuine, normal language works. Taken on the whole, these disparate elements combine in DDL as a route through active learning to effective, engaged grammar.

This study is sought to answer the following questions:

Q1: How effective is Data-Driven Learning (DDL) in improving the grammatical accuracy and lexical proficiency of Thai EFL learners compared to traditional teaching methods?

Q2: What effect does the use of Data-Driven Learning (DDL) have on learner engagement, autonomy, and satisfaction in Thai EFL classrooms?

Methodology

Participants

The research participants were 60 undergraduate students at the intermediate level studying in the EFL program at one of the most well-recognised universities in Thailand. The participants represented almost all faculties in the institution. These students were then randomly assigned to two groups: one for experiments-a group receiving instruction through approaches in Data-Driven Learning-and a control group, receiving traditional instruction. Furthermore, every precaution was taken so as to minimise bias selection and the effects that might ensure comparability between the two groups through the random assignments of the candidates to groups. Informed consent was attained prior to actual participation from the students. The consent process included a description of the purpose of the study, the procedures involved in the study, the potential benefits for the participants, and the voluntariness of their participation, with the freedom to withdraw from the study at any time or stage without any consequences. This was important in ensuring protection for ethical considerations and full disclosure.

Research design

The research adopted a quasi-experimental design using a pre-test/post-test structure with a control group in the test of effectiveness of DDL instruction. This design allows for robust comparisons of language proficiency improvement between the two groups. The pre-test/post-test model ensures that at baseline, the equivalence between the groups is measured, and that further comparisons can be made to show the impact of the intervention.

It was designed to isolate the effect of DDL strategies, with a control on the variables, as empirical evidence in effectively improving EFL learners' proficiency.

Procedure

The intervention lasted six weeks, and during that period, both groups received two 90-minute instructional sessions per week, amounting to 18 hours of teaching in total. The experimental group worked with language corpora and concordancing tools in search of authentic patterns of language, collocations, and grammar by working out data-driven exercises. These activities were targeted at promoting inductive learning, whereby the students themselves had to find the rules and patterns of the language; such an approach leads to much deeper absorption and long-term retention. On the other hand, the students in the control group went through a standard process with a textbook-based exercise program, focusing on mechanical learning and drills in the traditional manner. The instructional content was the same for both groups, therefore, the same topics were covered because they were aligned with the programme's learning outcomes. What is more, both groups completed a standardised language proficiency test validated for vocabulary and grammar on pre- and post-conditions. This pre-test/post-test framework allowed measurement of proficiency changes that were due to the instructional method.

Data collection

Data collection instruments were replicated to comprehensively establish the performance comparisons of the instructional methods applied. Data are based on the main sources: standardised tests before and after intervention in terms of vocabulary acquisition and grammatical competence. The test was prepared in such a way that it gave focus to ensure reliability and validity to meet the international standard for language testing. In addition to the proficiency tests, participants completed a post-intervention Likert-scale questionnaire to capture their perceptions of and attitudes toward the instructional approach. The questionnaire included items addressing such factors as engagement, perceived difficulty, and overall satisfaction with the learning experience. This triangulated approach to data collection provided both quantitative and qualitative insights into the outcomes of the study.

Data analysis

Data analysis included item-by-item data analysis to assure the accuracy and meaning of the results. The descriptive statistic calculations, including means and standard deviations, were used to summarise the pretest and post-test scores for the groups. Inferential statistical tests were used to test the significance of the changes observed. For testing the improvements from before to after in both groups, the paired sample t-tests were performed on the pretest and post-test scores. Independent sample t-tests were conducted on the improvement of mean score differences between the experimental and the control groups; hence, they allow for DDL instruction to be isolated as a cause of observed changes in these variables. Beyond this, effect size calculations were considered appropriate

for calculating practical significance. Lastly, in this regard, the data accruing from the Likert-scale questionnaires employed central tendency and dispersion measures, hence enabling an indication of attitude trends by participants.

Table 1.
Statistics for Pre-Test and Post-Test Scores

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pre_Test_Grammar	61.600	60	6.9828	.9015
	Post_Test_Grammar	74.825	60	11.1370	1.4378
Pair 2	Pre_Test_Vocabulary	55.150	60	6.1099	.7888
	Post_Test_Vocabulary	68.375	60	10.5756	1.3653

The paired samples statistics Table 1, reveal significant changes in participants' mean scores for grammar and vocabulary pre- and post-tests, suggesting a positive effect of the intervention.

Grammar

For grammar, the pre-test mean score was 61.60 (SD = 6.98), and the post-test mean score increased to 74.83 (SD = 11.14). The standard error of the mean (SEM) for the pre-test was 0.90, while for the post-test it was 1.44. This mean score increase of 13.23 indicates a substantial improvement in participants' grammar proficiency. The larger standard deviation in the post-test scores suggests increased variability among participants' performance, potentially reflecting varied engagement or individual learning rates.

Vocabulary

In vocabulary, the pre-test mean score was 55.15 (SD = 6.11), while the post-test mean rose to 68.38 (SD = 10.58). The SEM values were 0.79 for the pre-test and 1.37 for the post-test. The mean score improvement of 13.23 mirrors the gains observed in grammar. Similar to the grammar results, the post-test's higher standard deviation indicates greater variability in participants' vocabulary improvement.

Interpretation

The paired samples statistics indicate how intervention enhanced the ability of subjects in both grammar and vocabulary. As discussed above, the increased mean score gives an indication of the success of the intervention. Higher post-test standard deviation gives the indication of variation in individual learning outcomes. Small SEM gives evidence for the reliability of mean estimates to suggest that the improvements are statistically significant and not due to random chances. These findings constitute evidence of the effectiveness of the intervention in

advancing language proficiency, although further investigation is recommended into individual factors influencing variability.

Table 2.
Ranks for Engagement, Autonomy, and Satisfaction

	Group	Ranks		
		N	Mean Rank	Sum of Ranks
Engagement	1	30	45.50	1365.00
	2	30	15.50	465.00
	Total	60		
Autonomy	1	30	38.75	1162.50
	2	30	22.25	667.50
	Total	60		
Satisfaction	1	30	45.50	1365.00
	2	30	15.50	465.00
	Total	60		

Table 2 presents the ranks, giving a comparison of the mean ranks and sum of ranks for engagement, autonomy, and satisfaction for Group 1 (experimental group) and Group 2 (control group). The results show significant differences between the groups, indicating the influence of the experimental intervention.

Engagement

The mean rank for the experimental group, Group 1, is significantly higher at 45.50 compared to the control group, Group 2, which has a mean rank of 15.50. The sum of ranks for Group 1 is 1365.00, while Group 2's sum of ranks is 465.00. This huge difference in sum of ranks shows that the participants in the experimental group were much more engaged compared to the participants in the control group.

Autonomy

In terms of autonomy, Group 1 has the higher mean rank with 38.75, while Group 2 trails behind with a mean rank of 22.25. Overall, Group 1 participants had a sum of ranks of 1162.50, compared to Group 2, who totalled 667.50. Again, this would suggest that the intervention had a positive effect on persons' sense of autonomy; the difference between the two groups is not as large as in the case of engagement.

Satisfaction

Satisfaction reveals a similar picture to that of engagement, with the mean rank of 45.50 and sum of ranks of 1365.00 for Group 1 and mean rank of 15.50 and sum of ranks of 465.00 for Group 2. Such results indicate that the experimental group was significantly more satisfied than the control group.

Interpretation

Indeed, all the variables of engagement, autonomy, and satisfaction are outperforming the control group. A number of differences in average and sum of ranks testify for the efficiency of the intervention for such a boost of the attributes under research. Such a large difference in engagement and satisfaction does suggest that the intervention has been able to enhance participant experiences more than traditional approaches. Confirmation of these observed differences by means of statistical testing is required, using the Mann-Whitney U test.

Ethical considerations

Ethical considerations were strictly followed in the study. The purpose of the study, procedures involved, and possible benefits of the research were explained to the subjects in the informed consent. It was assured for them that the participation was totally voluntary and one could withdraw at any time without any adverse consequences. The confidentiality of all data was maintained by anonymising the data of participants during analysis and reporting. No personal data were linked to test results or survey responses. At the same time, as required for every research work, its due adherence to all the relevant regulations and ethical standards in educational research, the ethical approval was duly received from the institutional review board of the university. This method was followed with the purpose of rigorously assessing the effectiveness of DDL-based instruction with regards to improvement in language proficiency. Such a study, using a robust research design with detailed data collection and comprehensive statistical analysis, would provide empirical proof of the effectiveness of the DDL strategy in learning the English language as compared to the traditional method. The findings of this paper offer valued contributions to a set of best practices in the area of language learning improvements among EFL learners.

Discussion

These results make for strong evidence that DDL can work effectively as an instructional approach in developing grammar and vocabulary proficiency among EFL learners. Combined with participant feedback, the statistical analysis supports significant advantages of DDL in language education. This approach focused on the usage of corpus-based materials and concordancing tools to provide such authentic language exposure that would allow learners to come into direct contact with real linguistic patterns, thus enhancing deeper comprehension and practical application of language.

Pre- and post-test results confirm that the DDL intervention is effective. Indeed, the experimental group outperformed the control group in grammar and vocabulary proficiency, following traditional methods of instruction. For grammar, the experimental group's mean increased significantly from 61.60 to 74.83, while vocabulary scores increased from 55.15 to 68.38. These findings are supported by a range of theoretical positions on Second Language Acquisition, be it Krashen's Comprehensible Input Hypothesis or Schmidt's Noticing Hypothesis, since their findings highlight in detail how exposure to actual input enhances the process of language acquisition. On the contrary, DDL enabled students to both identify and internalise linguistic collocations autonomously, therefore performing better both for immediate performance but also long-term retention.

Quantitative data from the Likert-scale questionnaires further support the instructional benefits of DDL. According to the ranks analysis, the experimental group demonstrated much higher levels of engagement, autonomy, and satisfaction than the control group. For example, the mean rank for engagement in the experimental group was 45.50 compared to 15.50 in the control group. Similarly, autonomy and satisfaction scores were consistently higher for the DDL participants. The findings also bring out one even more important advantage of DDL as a methodology in inducing an attractive attitude change in the learner towards language learning; very key especially in EFL contexts, where most of the traditional methodologies are usually unable to sustain student motivation and active participation. Despite such promising results, the study also points to issues that arise when trying to apply DDL in the EFL classroom. The most significant issue seems to relate to the level of complexity that corpus-based materials may cause for lower proficiency learners. Indeed, as the methodology admits, the amount and difficulty of authentic corpus output can at times be beyond the cognitive abilities of students with a low foundational level of proficiency. In order to address this, the study used paper-based corpus materials as scaffolding to make the content accessible to learners of different proficiency levels. This is a reflection of the importance of adapting DDL resources to the diverse needs of students, which is crucial for maximising its effectiveness.

Other challenges include preparing and delivering the DDL materials themselves, which take a great deal of time in terms of work by teachers. For instance, the development itself of customised corpus-based tasks according to particular learning outcomes requires strong knowledge in both corpus linguistics and pedagogical perspectives. These findings underline a need for institutional support, including professional development for educators and collaborative efforts in material design. In this regard, DDL lacks the scalability to be a mainstream teaching method without proper training and resources.

Such findings of the study also bear on the wider implications for DDL and EFL education in Thailand. Due to structural differences, Thai learners generally face some specific linguistic problems while learning English.

Traditional methods, such as GTM, have proved quite inadequate to tackle such issues. This study has tried to present an alternative through which those gaps could be bridged by controlled exposure to high-frequency patterns and collocations through DDL. This approach also befits the learning outcomes of the program, making it relevant and practical in real-life communication contexts.

Most significantly perhaps, the authors further present it to note how DDL overcomes several notorious systemic barriers typical within the current Thai ELT provisions. Engaging activities conducted by Interactive Group work covering Corpus based focus encouraged efficient development both at a level of the students as Linguistic competencies while giving impetus for confident, independently learned results among learners thus reflecting an almost re-mapping from Teacher-oriented activities towards efficient concentrated activity engagement between learner-centred aspects showing inculcation and development consistent with more updated conceptual directions, therefore representing new kinds of experiential approaches being recognised to better philosophically drive more holistic learner-led, experience-focused perspectives to become updated at ground realities. Methodologically, the quasi-experimental design with a pre-test/post-test structure and a mixed-methods approach allows the study to ensure reliability and validity in its findings. The statistical analysis includes paired sample t-tests and ranks analysis, which are robust and thus provide strong support for the established causal relationship that DDL instruction improves language outcomes. Besides, quantitative and qualitative triangulation provides a comprehensive view of learners' experiences that enriches this study's contribution to the field.

The study raises some questions that could be useful in the future. For example, it would be useful to know how prior exposure to authentic use, individual learning styles, and the cultural context may influence the efficiency of DDL. In addition, further embedding DDL within incoming technologies, including artificial intelligence and adaptive learning platforms, is promising. Discussion of such questions may extend the use of DDL in different educational contexts for various groups of learners. The study was able to reflect the transformative potential of DDL in the EFL classroom, emphasising authenticity, learner autonomy, and active discovery where DDL overcomes many weaknesses in traditional grammar teaching. Certain problems still remain in the areas of access and teacher training; nonetheless, learner outcomes and classroom processes together suggest considerable advantages to the approach. The present study underlines the potential of innovative, data-driven pedagogies for responding to complex emerging needs within the language classroom. With the growing integration of educational technologies and learner-centred practices in institutions, DDL has proved to be one of the most promising frameworks for enhancing language proficiency and engaging EFL learners.

Conclusion

The study has unveiled the transformative potentials of DDL in EFL grammar instructions that are needed at the threshold in Thai education. Synthesising key principles from corpus linguistics and important learner-centred pedagogical ways, DDL presents an empowering alternative to the outdated traditional Grammar Translation Method (GTM) approaches. The findings bring into light how DDL can help resolve some long-standing issues in the teaching of grammar, such as the gap between grammatical knowledge and its application, learners' low level of engagement in learning, and the limitations related to structural differences between English and Thai.

This quasi-experimental design of the study reveals that DDL nurtures not only grammatical accuracy and lexical knowledge but also learner autonomy and critical thinking. The experimental group gained significantly higher in language proficiency compared to the control group, testifying to the effectiveness of DDL in promoting meaningful and lasting learning. These results are in line with SLA theories that give prominence to an active learner and the contextualisation of grammatical structures. By allowing students to investigate authentic patterns of language themselves through corpora, DDL bridges the gap between classroom learning and real-life language use, enabling learners to internalise grammar in a way that is both practical and memorable. The gains extend well beyond the linguistic, however, as the investigation also illustrates how DDL has the capacity to redefine class dynamics. Reported satisfaction and learner engagement in an experimental group must therefore be attributed to the inclusive and explorational nature of approach. Refocus away from passive role rule memorisations to active 'discovery brings about a return of the mantle of control toward the learner where DDL reharnesses individual learners to create their language progression. This is all the more relevant in the Thai EFL classroom, where traditional methods of teaching have normally led to students' passivity and teacher domination. The stress on collaboration and small-group activities inherent in DDL provides an effective means of overcoming some other challenges such as huge class sizes and unengaging content, thereby creating a warm and dynamic, inclusive classroom atmosphere.

These disadvantages again bring into focus that while the many benefits may accrue from the approach, one needs to weigh some practical challenges associated with DDL. This task could be formidable for low-proficiency learners not properly scaffolded for handling real data in corpora. It would also entail additional workload on the part of the teachers too to prepare the material for DDL. Again, this indicates focused training for the teachers as well as development of resource materials before introducing DDL to students. This requires institutions to invest in providing educators with the skills and tools they need to make full use of corpus-based methodologies. The exploitation of paper-based materials, together with guided exercises, are also ways to make DDL more accessible for a broader range of learners.

The implications of this study go beyond the Thai EFL context. In a world where language education is continually evolving with each tide of technological development, the insights gained from the present research inform the adoption of DDL across diverse educational settings. DDL provides a paradigmatic shift in teaching and learning grammar due to its priority on authenticity, learner autonomy, and active participation. Its correspondence to modern pedagogical principles and technological novelties places it in a unique position to solve the shortcomings of traditional methodology and improve language proficiency on an international scale.

This study underlines the possibility of DDL revolutionising grammar instruction in EFL contexts. Challenges apart, the merits of this approach are undisputed. While filling the gap between theory and practice, DDL additionally develops the grammatical competence of learners and turns the classroom into a space for active, meaningful, and enjoyable learning. While educators and policymakers are looking to innovate for the betterment of language education, DDL opens an exciting avenue toward the future of the educational environment in a more effective and learner-centred way.

Acknowledgement

The authors would like to express their deep appreciation to the students and faculty members who participated in this study for their time, effort, and valuable contributions. We also extend our gratitude to the university administration for helping to facilitate the research process and providing us with all the resources necessary in conducting this study efficiently. First and foremost, special thanks go to colleagues in the Department of English for their constructive feedback and encouragement in the course of this project. In addition, the authors would like to appreciate the effort of previous researchers in corpus linguistics and DDL whose efforts gradually paved the way for this study.

The authors have no conflict of interest to declare concerning this research. The study was performed independently, and no funding or any other form of influence was used, which guarantees the objectivity and integrity of the research. We hope that the results will be useful in current efforts toward innovative and better methodologies in language teaching within EFL contexts worldwide.

AI Acknowledgment

The authors declare that AI tools were used only for language proofing and refinement during the preparation of this manuscript. No such tools were used in generating ideas, data analysis, or in conceptual development to ensure integrity and originality.

Disclosure statement

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. No grant is involved in funding this study, hence there is no external financial influence. We consider that these conclusions will support or give continuity to those so far presented that contribute to innovations that are in development for effectively improving current methodologies used in teaching a new language to foreigners.

References

- Anyan, J. (2006). Different families, not distant cousins: Comparing Thai and English. *MED Magazine*.
- Biber, D., & Conrad, S. (2010). Corpus linguistics and grammar teaching. Retrieved from <http://www.pearsonlongman.com/ae/emac/newsletters/may-2010grammar.html>
- Boontam, P., & Phoocharoensil, S. (2018). Effectiveness of English preposition learning through data-driven learning. *3L: Language, Linguistics, Literature*, 24(3), 125–141.
- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3), 534–572. <https://doi.org/10.1111/j.1467-9922.2010.00566.x>
- Chambers, A. (2010). What is data-driven learning? In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 345–358). Routledge.
- Chingchit, O. (2008). Basic English grammar materials for EFL advanced beginners in Thailand.
- Choomthong, D. (2014). Preparing Thai students’ English for the ASEAN economic community: Some pedagogical implications and trends. *Language Education and Acquisition Research Network (LEARN) Journal*, 7(1), 45–57.
- Evison, J. (2010). What are the basics of analysing a corpus? In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 122–135). Routledge.
- Flowerdew, L. (2012). *Corpora and language education*. Palgrave Macmillan.
- Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359–370). Routledge.
- Huang, Z. (2014). The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL*, 26(2), 163–183. <https://doi.org/10.1017/S0958344014000020>
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *ELR Journal*, 4, 1–16.
- Jones, C., & Waller, D. (2015). *Corpus linguistics for grammar: A guide for research*. Routledge.
- Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh University Press.
- Nugraha, S. I., Miftakh, F., & Wachyudi, K. (2017). Teaching grammar through Data-Driven Learning (DDL) approach. *Advances in Social Science, Education and Humanities Research*, 125, 300–303.
- Nunan, D., & Bailey, K. M. (2009). *Exploring second language classroom research: A comprehensive guide*. Heinle.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- Römer, U. (2008). Corpora and language teaching. In A. Ludeling & M. Kyto (Eds.), *Corpus linguistics: An international handbook* (pp. 112–130). De Gruyter.

- Sayadi, Z. A. (2007). *An investigation into first year engineering students' oral classroom participation: A case study*. University of Technology.
- Sclove, S. L. (2001). Notes on Likert scales.
- Sinclair, J. (1990). *Collins COBUILD: English grammar*. Collins.
- Sinclair, J. (1991). *Collins COBUILD: English grammar*. HarperCollins Publishers.
- Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL*, 26(2), 184–201.
<https://doi.org/10.1017/S0958344014000081>
- Sripicharn, P. (2002). Introducing data-driven learning: Theory, practice, and evaluation. *Journal of Liberal Arts*, 2(2), 100–111.
- Sripicharn, P. (2010). How can we prepare learners for using language corpora? In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 371–384). Routledge.
- Taylor, J. R. (2002). *Cognitive grammar*. Oxford University Press.
- Yepes, G. R., & Krishnamurthy, R. (2010). Corpus linguistics and second language acquisition: The use of ACORN in the teaching of Spanish grammar. *Living Languages*, 55(1).
- Yunus, K., & Awab, S. (2012). The effects of the use of module-based concordance materials and data-driven learning (DDL) approach in enhancing the knowledge of collocations of prepositions among Malaysian undergraduate law students. *The International Journal of Learning*, 18(9), 165–181.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution.