JOURNAL OF RESEARCH STUDIES IN ENGLISH LANGUAGE TEACHING AND LEARNING

To cite this article in APA 7th style:

Vedantham, R. (2025). The influence of data-driven learning on learner autonomy and vocabulary acquisition. *Research Studies in English Language Teaching and Learning*, *3*(2), 398–410. https://doi.org/10.62583/rseltl.v3i2.83

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.

ISSN (online): 2977-0394

KEYWORDS
data-driven learning (DDL), learner autonomy, vocabulary acquisition, corpus linguistics, EFL instruction.

The influence of data-driven learning on learner autonomy and vocabulary acquisition

Raghunath Vedantham[1]
[1]*Chaudhary Charan Singh University, Meerut, Uttar Pradesh, India*

Abstract

This study examines the influence of Data-Driven Learning (DDL) on learner autonomy and vocabulary gain in English as a Foreign Language (EFL) classroom. Founded on corpus linguistics, DDL enables learners to handle authentic language data, fostering inductive learning and autonomous exploration. While earlier research has testified to the usefulness of corpus-based teaching in enhancing lexical awareness, few empirical studies have examined its impact on learner autonomy and vocabulary gain in combination. This study follows a quasi-experimental research design with undergraduate EFL students in India, employing integrated quantitative and qualitative data collection procedures. Learners utilised corpus tools to explore word patterns, collocations, and grammatical patterns, facilitating independent learning. The findings demonstrate that DDL enhances vocabulary retention through contextualised linguistic input and fosters autonomy via independent corpus exploration. However, corpus tool navigation and data interpretation concerns indicate the need for structured instructional support. The study contributes to the literature by offering empirical testimony on the dual benefits of DDL, underlining its potential in fostering active learning and long-term vocabulary development. Pedagogical implications of integrating corpus-based teaching in language curricula are discussed, providing insights for language teachers interested in supplementing EFL teaching through DDL.

Introduction

Among the various pedagogical approaches introduced through technological advancements in language instruction, Data-Driven Learning (DDL) has proven effective in promoting learner autonomy and vocabulary acquisition. in promoting learner autonomy and vocabulary acquisition. DDL, grounded in corpus linguistics, enables learners to investigate actual language usage through corpus-based tools, thereby fostering an inductive language learning process (Johns & King, 1991). This shift towards learner-centred instruction is aligned with broader educational trends endorsing constructivist and experiential learning paradigms (Boulton, 2010). With increased accessibility to digital resources and corpus tools, the need to investigate the pedagogical value of DDL in English as an EFL context has grown exponentially. Learner autonomy, or the ability to take charge of one's own learning, has long been recognised as a central factor in language learning (Holec, 1981). Traditionally delivered, teacher-centred language courses have proven ineffective in fostering independent learning abilities, replicating instead a model of passive knowledge reception instead of active language construction (Little, 1991). In contrast, DDL positions the student to explore linguistic patterns, identify collocations, and infer grammatical structures through direct exposure to language data (Huang, 2011; Yoon, 2008). Not only does this autonomous learning process reinforce vocabulary retention, but it also enhances deeper cognitive engagement, as learners become active participants in their linguistic development (Boulton, 2012). Although its theoretical advantages are evident, the extent to which DDL facilitates learner autonomy in actual classroom practice is comparatively understudied, and in the EFL context, this is particularly the case.

Vocabulary acquisition is another core issue in second language learning, as it has direct implications for communicative competence and academic success (Nation, 2001). Traditional vocabulary instruction relies on rote memorisation, which is unlikely to foster long-term retention or contextual understanding (Schmitt, 2000). DDL offers an alternative by exposing learners to authentic linguistic data, allowing them to see word pattern applications and collocational behaviour in real contexts (Frankenberg-Garcia, 2012). Studies have demonstrated the potential of corpus-based approaches in improving vocabulary knowledge (Yılmaz & Soruç, 2015; Lin & Lee, 2015), yet empirical studies remain limited regarding their impact on fostering long-term vocabulary development in EFL learners. In spite of the growing body of research on DDL, several gaps remain. Although studies have emphasised its benefits on lexical awareness and grammatical accuracy, relatively few have investigated its impacts on learner autonomy and vocabulary acquisition in tandem (Rezaee et al., 2014). In addition, concerns regarding students' ability to use corpus tools independently and accurately interpret concordance data have been raised (Kennedy & Miceli, 2001). These issues underscore the need for additional research to determine optimal instructional approaches to using DDL effectively in diverse educational contexts.

Literature review

As educational technology was integrated into language teaching, different pedagogical approaches and technological applications have abounded, each with its own set of terminologies such as computer-assisted language learning (CALL) (Soruç, 2015) and multimedia learning (Mayer, 2001, 2005). Others include e-learning (Peng, Su, Chou, & Tsai, 2009) and, with the advent of mobile technologies, m-learning, which employs mobile technologies such as personal digital assistants (PDAs), mobile phones, and laptops (Hockly, 2013; Şad, 2008; Şad & Göktaş, 2014; Saran, et al., 2008; Saran, et al., 2009). In addition, corpus-assisted

language learning has been a point of interest (Aston, et al. 2004; Çelik & Elkatmış, 2013; Huang, 2011). These approaches have revolutionised the effectiveness of language learning to a great extent, but corpus-assisted language learning, or data-driven learning (DDL), remains a research interest. The difference between using concordance lines in exploring word meaning and traditional vocabulary teaching in second or foreign language learning has drawn a lot of scholarly interest.

Historically, corpus linguistics has evolved from its beginnings in scepticism when the first computer corpus, the Brown Corpus, was deemed "a useless and foolhardy enterprise" (Francis, 1992, p. 28) to becoming an indispensable tool that has "revolutionised" dictionary-making processes (O'Keeffe, et al., 2007, p. 21). Corpus-supported learning can be employed to investigate the study of polysemy, semantic prosody, phraseology, and true grammar (Conrad, 2000; Dönük, 2016; Kılıçkaya, 2015; Reppen, 2010; Uysal, et al., 2013). Corpus-based English reading courses for academic purposes have also been utilised by teachers (Kırkgöz, 2006). DDL, the principal pedagogical approach to the utilisation of corpora in language instruction, has been defined as "the use in the classroom of computer-generated concordances to get students to explore the regularities of patterning in the target language" (Johns & King, 1991, p. iii) and later redefined as "the attempt to cut out the middleman as far as possible and to give direct access to the data" (Johns, 1994, p. 297). This approach encourages learner autonomy, enabling students to discover word use and collocational tendencies independently (Çelik, 2011; Huang, 2011).

DDL is underpinned cognitively by linguistic theory, with learners engaging in pattern recognition to identify structural regularities through an inductive process (Geluso & Yamaguchi, 2014, p. 227). The method is underpinned by psycholinguistic principles, with learners engaging in "psycholinguistic guessing games" through the utilisation of concordance lines (Yılmaz & Soruç, 2015, p. 2628) and using corpora as a "mediational tool" (Vygotsky, 1978). Schmidt's (1994) "noticing hypothesis" also supports the method, in that learners need to consciously notice linguistic features for acquisition to occur. Boulton (2010) posits that DDL is not based on explicit grammar instruction but instead encourages learners to deduce linguistic structures from corpus data, with corpora being used as an "awareness-raising tool." Despite its advantages, DDL is not without criticism. Kennedy and Miceli (2001) reported that some students were demotivated by DDL-based tasks, whereas Rezaee, et al. (2014) got positive feedback from students working on such tasks. DDL has been praised for providing opportunities for authentic input (Johns, 1991; Sun & Wang, 2003; Yoon, 2011), building learner autonomy (Huang, 2011; Lin & Lee, 2015; Starfield, 2004; Yoon, 2008), and promoting active engagement. However, Boulton (2010) is of the opinion that DDL is not effective for teaching certain grammar points, though research has proved its potential in the correction of L2 grammatical errors (Frankenberg-Garcia, 2012; Gaskell & Cobb, 2004; Gilmore, 2009; Quinn, 2014). In addition, Rezaee et al. (2014) and Smart (2014) are of the opinion that DDL can raise learners' awareness of collocations and grammar through exposure to multiple contextual examples (Thurstun & Candlin, 1998; Wu, Witten, & Franken, 2010).

Empirical studies have provided strong support for DDL. Thurstun and Candlin (1998) developed concordance-based materials, finding that learners appreciated the innovative approach despite initial difficulties with truncated concordance lines. Yoon (2008) demonstrated that corpus use increased students' linguistic awareness, influencing writing processes and collocational proficiency. Çelik and Keser (2010)

found a positive correlation between online corpus consultation and vocabulary acquisition, while Çelik (2011) reported that DDL facilitated vocabulary retention more than online dictionary consultation. Frankenberg-Garcia (2012) demonstrated that corpus examples were more effective than dictionary definitions in allowing learners to produce syntactically correct sentences. Some studies have yielded counterintuitive findings. Ünaldı, et al. (2013) found decontextualised vocabulary instruction to outperform corpus-based instruction, although methodological flaws, e.g., the use of multiple-choice testing, may have influenced the result. Geluso and Yamaguchi (2014) reported Japanese university students to have some reservations about corpus use but acknowledged its utility for vocabulary learning. Özdemir (2014) reported medical students' preference for corpus-based vocabulary learning, demonstrating its feasibility in ESP contexts. Further research has continued to corroborate the effectiveness of DDL. Yılmaz and Soruç (2015) found that Turkish EFL students who participated in corpus-based vocabulary learning outperformed a control group, finding DDL enjoyable and empowering. Uçar and Yükselir (2015) found that students who used a corpus had a more nuanced understanding of verb-noun collocations. Lin and Lee (2015) found that Taiwanese instructors preferred DDL to traditional grammar translation, describing the former as innovative and engaging. Similarly, Özbay and Kayaoğlu (2015) found that Turkish EFL instructors, after undergoing corpus training, found corpus tools to be helpful for language exploration.

Recent qualitative research has provided more insight. Tekin and Soruç (2016) found that international high school students described corpus-based tasks as "easy," "fun," and "practical," though sometimes complicated. Aşık et al. (2016) found that Turkish ELT students gained greater lexical awareness from DDL-based tasks. These findings align with more general studies demonstrating that DDL facilitates deeper lexical processing (Frankenberg-Garcia, 2014; Leel, 2011). Despite the proven effectiveness of DDL in fostering independent learning through corpus use, Mukherjee (2006) found a persisting mismatch between corpus research and classroom practice. The reluctance of language teachers toward DDL, as proven by Conrad (2005) and Flowerdew (2012), underlines the need for greater awareness and training in corpus methodologies. Future studies should explore DDL's application to varied learner groups and proficiency levels in order to strengthen its pedagogical implications. The present study aims to fill this gap by providing answers for the following research questions:

Q1: How does the implementation of Data-Driven Learning (DDL) impact vocabulary acquisition among undergraduate EFL learners?

Q2: To what extent does Data-Driven Learning (DDL) promote learner autonomy in English language learning?

Methodology

Participants

The research sample consisted of sixty undergraduates studying in an English language course at a university in Meerut, Uttar Pradesh, India. The sample included students from the first year of various academic fields such as engineering, business studies, and social sciences, and with varying levels of English proficiency. The selection criterion included students with at least a B1 level in the Common European Framework of Reference for Languages (CEFR) according to the results of the placement test. Also excluded were students who had attended a formal English language course outside the curriculum within the last year so that the

gains could be assumed to be an outcome of the intervention itself. Participants were selected by convenience sampling, as they were already enrolled in the researcher's classes. In an effort to enhance representativeness, an effort was made to recruit students from diverse academic backgrounds and linguistic experiences. The research ensured equal representation of male and female students, with thirty-two females and twenty-eight males. The age of the participants ranged from eighteen to twenty-one years old, with the mean age being nineteen. The majority of the participants were native speakers of Hindi, while a lower percentage were bilingual or multilingual, with proficiency in regional languages including Punjabi and Bengali.

Research design

This study employed a quasi-experimental within-subjects design to examine the effects of DDL on learner autonomy and vocabulary gain. A within-subjects design was adopted because it allows each participant to act as his or her own control, thereby reducing variability and increasing the accuracy of pre-test and post-test comparison. The design is particularly well-suited for interventions whose aim is to quantify learning gains over time. The study took a pre-test/post-test design, whereby the participants' vocabulary knowledge and autonomy levels were tested prior to and following the intervention. The independent variable in this study was the DDL-based instruction, while the dependent variables were vocabulary gain and learner autonomy, which were quantified using validated pre-test and post-test measures. The pre-test was administered to establish a baseline of the vocabulary knowledge and autonomy of the participants in language learning. The post-test was administered after the intervention to determine the degree of improvement in both areas. The design controlled confounding variables by maintaining the learning environment constant, and all students had equal access to learning materials and facilities. Furthermore, the study used a mixed-methods design with both quantitative and qualitative data collection methods to attain an in-depth insight into the effect of DDL on learners. The quantitative strand measured statistical differences in scores, while the qualitative strand examined students' experiences, perceptions, and interaction with DDL tools. Blending the methods supported the internal validity of the study and the credibility of results.

Procedure

Two baseline measures were taken from participants before the intervention. One was a vocabulary test to measure their knowledge of target academic and general English vocabulary. The other was an autonomy questionnaire, modified from previously validated measures, measuring participants' self-reported language learning autonomy. An orientation session was set up to familiarise students with Data-Driven Learning, providing an overview of corpus-based language learning. Students were familiarised with corpus tools such as COCA, BNC, and Sketch Engine in the session and practiced extracting and analysing authentic language use from the tools. The intervention lasted for six weeks, with two ninety-minute sessions per week. Sessions followed a structured sequence to facilitate progressive learning. Students were introduced to corpus tools in the first two weeks and explored concordance lines, collocations, and frequency lists. Tasks included identifying common patterns in word use and differentiating between synonyms through contextual evidence. In weeks three and four, students carried out guided exercises in which they formulated their own rules based on corpus evidence. They worked in small groups, comparing findings and contrasting them with traditional dictionary definitions. The final two weeks of the intervention were independent corpus research projects, where students carried out autonomous language investigation. They documented the research process and reflections on autonomy in learning through learning logs.

After the intervention, the students completed the same vocabulary test and autonomy questionnaire used in the pre-test session. A qualitative reflection questionnaire was given to gather data on students' experience of working with DDL tools. This was followed by a debriefing session where, through a focus group discussion, students had a chance to comment on the contribution of DDL to autonomy and vocabulary development. The qualitative data presented through the reflection questionnaires and focus group discussions complemented the quantitative test scores, offering deeper insight into students' learning experience.

Materials

The study used a variety of learning support and assessment resources to facilitate an in-depth examination of the students' learning and interaction with DDL. Corpus-based tools, including the Corpus of Contemporary American English (CCAE), the British National Corpus (BNC), Sketch Engine, and AntConc, were used to facilitate language investigation and exploration. The tools allowed the students to learn from actual use of language, find collocations, explore word frequency, and conduct concordance-based learning activities. The vocabulary test consisted of fifty multiple-choice items for measuring the knowledge of high-frequency words, academic words, and the identification of collocational patterns among the students. The test was developed based on corpus data for ensuring a reference to authentic language use. The items were reviewed for their suitability and level of difficulty by three linguistic experts, and a pilot test was also administered on another group of students for clarifying ambiguous questions before final administration. To measure learner autonomy, the research employed a validated Likert-scale questionnaire that assessed students' ability to set learning goals, monitor progress, and employ independent learning strategies. The questionnaire had twenty-five items across self-regulation, resourcefulness, and motivation dimensions in language learning. It was adapted from available validated measures and reviewed by educational psychologists and language learning specialists for greater reliability and applicability in an EFL context. Besides structured measures, participants maintained comprehensive learning logs during the intervention. The logs captured students' utilisation of corpus tools, reflections on language discovery, and difficulties in autonomous learning. The logs provided qualitative information on how students utilised DDL and how their learning strategies evolved over time. The log data were later analysed using thematic coding to identify patterns in autonomous learning behaviours.

To solicit students' perspectives on the effectiveness of the intervention, an open-ended reflective survey was administered at the end of the study. The survey prompted students to describe their experience of working with corpus tools, the strengths and weaknesses they encountered, and the perceived impact on their language learning autonomy and vocabulary gain. With a small group of participants, a focus group discussion was also conducted to provide an opportunity for in-depth discussion of their learning experience. The researcher facilitated the focus group, following a semi-structured guide to prompt participants to discuss the key themes of learner autonomy and vocabulary gain. The discussions were audio-recorded, transcribed, and analysed to obtain qualitative feedback on students' interaction with DDL and their overall perceptions of its effectiveness.

Data analysis

Data analysis was quantitative and qualitative in nature. Repeated measures ANOVA was employed to compare pre-test and post-test scores on autonomy and vocabulary acquisition. Pillai's Trace, Wilks' Lambda,

Hotelling's Trace, and Roy's Largest Root were employed to determine effect sizes. Descriptive statistics in the form of standard deviations and means were reported to provide a comprehensive overview of the results. Thematic analysis was employed on focus group discussions and student reflections, coding being conducted with the assistance of NVivo software to identify the significant themes related to interaction with DDL tools and learner autonomy.

Reliability and validity

For reliability, the internal consistency of the autonomy questionnaire and vocabulary test was computed with Cronbach's alpha, which was higher than 0.80, indicating high reliability. This measure helped ensure that the questionnaire and test consistently measured the participants' vocabulary knowledge and levels of autonomy reliably. Test-retest reliability was also demonstrated using a subsample of the participants who took the tests again after a two-week interval, yielding a high correlation, which further demonstrated the reliability of the tools. For validity, content validity was ensured by expert review in which three EFL experts reviewed the test items and survey questions for relevance and appropriateness to the study objectives. Construct validity was checked using factor analysis, which helped confirm that the questionnaire was indeed measuring what was intended to measure in terms of learner autonomy. Criterion-related validity was ensured by correlating participants' performance in the vocabulary test with their prior English proficiency scores, which demonstrated a significant correlation between the measures. To enhance the strength of findings, triangulation was employed in the study by mixing various sources of data like quantitative test scores, qualitative reflections, and observational data. The research design reduced the potential for bias and added to the credibility of findings. Further, all the tests were administered under standardised conditions to facilitate consistency in testing procedures. Pilot testing of the instruments was conducted with a different group of students in an attempt to identify and refine ambiguous items before large-scale administration.

Ethical considerations

There were research ethics that were strictly followed. It was a voluntary response, and the students were told that they could withdraw at any point without any penalty. Confidentiality was maintained, and there was no personally identifiable information revealed. Storage of data was in accordance with institutional ethical principles, with anonymised responses being stored securely. These served to ensure that the research was ethically carried out while safeguarding the rights and privacy of research participants. The above procedure is a rigorous protocol for examining the effectiveness of Data-Driven Learning for facilitating learner autonomy and vocabulary acquisition in EFL contexts. All participants were briefed on the study objective, and they signed an informed consent before they took part in the study. Ethical clearance was sought and obtained from the university Institutional Review Board (IRB) prior to data collection. A confidentiality agreement was also signed by the researcher to ensure that all data would be anonymised and that participant

information would not be disclosed to third parties. Participants were provided with a detailed briefing session explaining the voluntary nature of the study, the right to withdraw at any time, and the possible risks and benefits of taking part.

Analysis

The results of the Multivariate Tests for Within-Subjects Effects in Table 2 indicate a statistically significant improvement in vocabulary scores from the pre-test to the post-test as shown in Table 1.

Table 1

Within-Subjects Factors

Measure: MEASURE_1

Time	Dependent Variable
1	Vocab_PreTest
2	Vocab_PostTest

Table 2

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
Time	Pillai's Trace	.735	163.644[b]	1.000	59.000	.000
	Wilks' Lambda	.265	163.644[b]	1.000	59.000	.000
	Hotelling's Trace	2.774	163.644[b]	1.000	59.000	.000
	Roy's Largest Root	2.774	163.644[b]	1.000	59.000	.000

a. Design: Intercept
Within Subjects Design: Time
b. Exact statistic

Pillai's trace showed a large effect size, V * = 0.735, F(1, 59) = 163.644, p < .001, indicating that 73.5% of the variance in vocabulary performance is accounted for by the time effect. Wilks' lambda also showed a significant effect,* Λ * = 0.265, F(1, 59) = 163.644, p < .001, demonstrating that only 26.5% of the variance is not accounted for, further confirming the strong effect of the intervention. Similarly, Hotelling's trace (T = 2.774, p < .001) and Roy's largest root (R = 2.774, p < .001) showed that the pre-test to post-test change in* Table 1 was substantial. Taken together, these findings in Table 2 provide evidence of a large improvement in vocabulary knowledge due to the intervention, with a strong within-subjects time effect. The large effect size suggests that the instructional approach in this research had a substantial and practical impact on vocabulary learning.

Table 3

Within-Subjects Factors

Measure: MEASURE_1

Time	Dependent Variable
1	Autonomy_PreTest

2	Autonomy_PostTest

The Multivariate Tests for Within-Subjects Effects in Table 4 indicate a statistically significant improvement in autonomy scores from the pre-test to the post-test as shown in Table 3.

Table 4

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
Time	Pillai's Trace	.535	67.777[b]	1.000	59.000	.000
	Wilks' Lambda	.465	67.777[b]	1.000	59.000	.000
	Hotelling's Trace	1.149	67.777[b]	1.000	59.000	.000
	Roy's Largest Root	1.149	67.777[b]	1.000	59.000	.000

a. Design: Intercept
Within Subjects Design: Time
b. Exact statistic

Pillai's trace revealed a moderate effect size, V * = 0.535, F(1, 59) = 67.777, p < .001, which means that 53.5% of the variance in autonomy scores is accounted for by the time effect. Wilks' lambda also revealed a significant effect,* Λ * = 0.465, F(1, 59) = 67.777, p < .001, indicating that 46.5% of the variance is not accounted for, though the autonomy increase is still substantial. Hotelling's trace (T = 1.149, p < .001) and Roy's largest root (R = 1.149, p < .001) also showed that the change in autonomy scores from pre-test to post-test in* Table 3 was significant. Cumulatively, these findings in Table 4 reveal that the intervention had a statistically significant and large effect on autonomy development, with a moderate to strong within-subjects time effect. Although the effect size is relatively smaller compared to that achieved for vocabulary, the findings suggest that the instructional approach made a positive contribution to fostering learner autonomy.

Discussion

The findings of this study provide robust empirical support for the positive impact of DDL on vocabulary acquisition and learner autonomy in EFL classrooms. The results, indicating statistically significant improvement in both vocabulary knowledge and autonomy, align with earlier studies that have explored the pedagogical value of corpus-assisted language learning (Aston, et al., 2004; Huang, 2011; Yoon, 2008). The subsequent discussion integrates the findings in the light of current literature and delineates the implications of DDL for language learning and teaching.

Vocabulary Acquisition through Data-Driven Learning

The significant increase in vocabulary scores from pre-test to post-test is evidence that DDL is an effective way of building lexical knowledge in EFL learners. The findings corroborate the results of previous studies that have reported the effectiveness of corpus-based approaches to vocabulary learning (Çelik & Keser, 2010; Frankenberg-Garcia, 2012; Yılmaz & Soruç, 2015). The fact that students were able to find patterns of word

usage, differentiate between synonyms, and get a feel for collocations illustrates that corpus tools offer a more authentic and context-rich learning environment than the conventional dictionary-based approach (Thurstun & Candlin, 1998; Wu et al., 2010). Corpus-based learning encourages learners to engage with real linguistic data, making vocabulary learning more meaningful and pertinent. As opposed to decontextualised lists of words or rote memorisation, DDL exposes learners to words in numerous contexts, allowing for deeper lexical processing (Frankenberg-Garcia, 2014; Leel, 2011). The substantial effect size observed in this study ($V = 0.735$, $F(1, 59) = 163.644$, $p < .001$) underscores the strong influence of DDL on vocabulary acquisition, corroborating findings from prior studies that have demonstrated the effectiveness of corpus consultation in enhancing learners' ability to internalise new words (Yoon, 2008; Özdemir, 2014).

One of the reasons for these results is the inductive nature of corpus-based learning, which aligns with cognitive linguistic principles and the noticing hypothesis (Schmidt, 1994). By engaging with corpus data, the learners are compelled to actively discover linguistic patterns, rather than passively being instructed. Such active learning not only fosters increased retention but also the application of newly acquired vocabulary in context (Geluso & Yamaguchi, 2014). Furthermore, collaborative discussion and peer interaction made possible through the corpus-based tasks may have contributed towards reinforcing vocabulary acquisition, as accounted for in Vygotsky's (1978) sociocultural theory of learning.

Development of Learner Autonomy

The results of this study also indicate a statistically significant improvement in learner autonomy, supporting previous research that has highlighted the role of DDL in fostering independent learning (Huang, 2011; Lin & Lee, 2015; Yoon, 2008). The moderate effect size ($V = 0.535$, $F(1, 59) = 67.777$, $p < .001$) suggests that while the influence of DDL on autonomy is substantial, it is slightly less pronounced than its effect on vocabulary acquisition. This finding aligns with studies that have noted variability in learners' ability to navigate corpus tools independently (Kennedy & Miceli, 2001; Conrad, 2005).

One of the primary reasons for the observed increase in autonomy is the independent nature of DDL tasks. Students were required to search corpus data independently, identify patterns, and draw conclusions about word usage. This encouraged them to be responsible for their own learning, in accordance with Boulton's (2010) argument that corpus tools are an awareness-raising as opposed to an explicit instructional tool. The autonomy questionnaire results, coupled with students' reflections and learning logs, suggest that many students became increasingly confident in their ability to self-regulate their learning and engage in independent vocabulary enrichment.

Despite such encouraging outcomes, it is important to note challenges in employing DDL to facilitate autonomy. Students did report some initial difficulties with the use of corpus tools and interpreting concordance lines, a problem echoed in earlier research as well (Rezaee et al., 2014; Smart, 2014). Such difficulties may have potentially contributed to the relatively lower effect size for autonomy compared to vocabulary acquisition. However, the scaffolded nature of the intervention, where guided corpus use was followed by independent work, likely mitigated such difficulties and enabled gradual learner adaptation to DDL methods.

Pedagogical Implications

The findings of this study have important implications for language teachers and curriculum planners who wish to introduce corpus-based instruction into EFL classrooms. Firstly, the results suggest that DDL can serve as a useful supplement to traditional vocabulary instruction, particularly in learning contexts where contact with authentic language is limited. The use of corpus tools enables learners to engage with naturally occurring language data and, in doing so, bridges the gap between language learning in the classroom and language use in the real world (Flowerdew, 2012; Mukherjee, 2006).

Second, the findings underscore the need for teacher training in corpus methodologies. As Conrad (2005) and Flowerdew (2012) note, one of the greatest obstacles to the adoption of corpus-based instruction is that instructors remain largely unfamiliar with the tools and their pedagogical applications. Providing instructors with professional development workshops in which they learn about the benefits and pragmatics of DDL may be what is needed to promote wider adoption and ensure that students are adequately supported in learning to use corpus tools. In addition, as much as the research proves the efficacy of DDL for the development of learner autonomy, it also emphasises the role of scaffolding. Students might first need to be given structured support in using corpus tools before they can on their own use them independently. A gradual shift from teacher-centred to student-centred corpus exploration can be used to pre-empt possible challenges and make learning even smoother (Tekin & Soruç, 2016).

Limitations and future research

Although this research provides robust evidence of DDL effectiveness, several limitations need to be mentioned. Most significantly, the research was conducted at one institutional context with students of a single demographic profile. Future research needs to examine the implementation of DDL with various learner groups and proficiency levels to determine whether similar gains are obtained in various instructional contexts. Moreover, while the quantitative findings demonstrate significant improvement in vocabulary acquisition and autonomy, more qualitative research can provide further insight into learners' DDL experiences. Longitudinal studies following learners' corpus tool use over a considerable period of time can provide enlightening information on the long-term impact of DDL on language learning outcomes. Future

research must examine the effectiveness of different corpus tools and instructional designs with the goal of determining optimum approaches to the integration of DDL into language curricula. Research into the potential for hybrid approaches to learning that combine corpus-based learning and other approaches to instruction can provide further insight into maximising the potential of DDL in EFL contexts.

Conclusion

This study provides enough evidence for the effectiveness of Data-Driven Learning in EFL teaching, justifying its crucial contribution to enhancing vocabulary development and building learner autonomy. The results suggest that corpus-based learning may promote deeper lexical processing and long-term retention, as well as enabling self-directed learning. The significant development of vocabulary acts as proof of the effectiveness of DDL in enabling students to notice actual language use, develop collocational awareness, and build linguistic intuition. Moreover, the formation of learner autonomy aligns with the findings of previous research, securing the place of DDL in enabling students to take responsibility for their own learning.

Although it is beneficial, some of the problems associated with corpus navigation and interpretation still persist, necessitating planned scaffolding and progressive familiarisation with corpus methods. Addressing such problems through instructional guidance and teacher training can enhance the utility of DDL in language teaching even further. Future research must follow up on the wider application of DDL in various learning contexts, proficiency levels, and learning outcomes, while investigating hybrid models of teaching using corpus-based learning in conjunction with other pedagogical approaches. These results add to the growing research literature in support of corpus-assisted language learning and point to the need for future research into the long-term impact of EFL learners.

Acknowledgment

I would like to express my gratitude to all the participants who spent their time and effort on this research. Their involvement and contribution were invaluable in exploring the use of Data-Driven Learning (DDL) in the English language classroom. I would also like to thank my mentors and colleagues for their constructive feedback and support throughout this research process.

I conducted this study independently, and I hereby state that no funds were received from any organisation or institution. Furthermore, there are no conflicts of interest in this study.

AI Acknowledgment

I acknowledge the use of ChatGPT (https://openai.com/chatgpt) for language and grammar checking only. These prompts entail requesting grammatical correctness, clarity, and coherence at the sentence level. The output from these prompts was used for refining language use for grammatical correctness and readability without altering the original content or arguments. While the author acknowledges the use of AI, the author

attests that he is the sole author of this article and take full responsibility for the contents herein, as per COPE guidelines.

References

Aşık, A., Vural, A. S., & Akpınar, K. D. (2016). Lexical awareness and development through data-driven learning: Attitudes and beliefs of EFL learners. *Journal of Education and Training Studies, 4*(3), 87–96.

Aston, G., Bernardini, S., & Stewart, D. (Eds.). (2004). *Corpora and language learners (Vol. 17).* Philadelphia, PA: John Benjamins Publishing.

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

Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. *Revista de Lenguas para Fines Específicos, 18*, 5–20.

Bozpolat, E. (2016). Investigation of the self-regulated learning strategies of students from the faculty of education using ordinal logistic regression analysis. *Educational Sciences: Theory & Practice, 16*, 301–318. https://doi.org/10.12738/estp.2016.1.0281

Çelik, S. (2011). Developing collocational competence through web-based concordance activities. *Novitas-ROYAL (Research on Youth and Language, 5*(2), 273–286.

Çelik, S., & Keser, H. (2010). The correlation between learners' logs of navigations through online corpora and lexical competence. *Ankara Üniversitesi Egitim Bilimleri Fakultesi Dergisi, 43*(2), 149–170.

Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly, 34*(3), 548–560.

Conrad, S. (2005). Corpus linguistics and L2 teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 393–409). Mahwah, NJ: Lawrence Erlbaum.

Crabbe, D. (1993). Fostering autonomy from within the classroom: The teacher's responsibility. *System, 21*(4), 443–452.

Flowerdew, L. (2012). *Corpora and language education.* New York, NY: Palgrave Macmillan.

Francis, W. N. (1992). Language corpora B.C. *Directions in Corpus Linguistics: Proceedings of Nobel Symposium, 82*, 17–32.

Frankenberg-Garcia, A. (2012). Learners' use of corpus examples. *International Journal of Lexicography, 25*(3), 273–296. https://doi.org/10.1093/ijl/ecs011

Frankenberg-Garcia, A. (2012). Learners' use of corpus examples. *International Journal of Lexicography, 25*(3), 273–296.

Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL, 26*(2), 128–146.

Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL, 26*(2), 225–242.

Hockly, N. (2013). Interactive whiteboards. *ELT Journal, 67*(3), 354–358.

Huang, L. S. (2011). Corpus-aided language learning. *ELT Journal, 65*(4), 1–4. http://dx.doi.org/10.1093/elt/ccr031

Johns, T. F. (2002). Data-driven learning: The perpetual challenge. *Language and Computers, 42*(1), 107–117.

Johns, T. F., & King, P. (1991). Classroom concordancing. *English Language Research Journal, 4*, 27–45.

Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology, 5*(3), 77–90.

Lin, M.-H., & Lee, J.-Y. (2015). Data-driven learning: Changing the teaching of grammar in EFL classes. *ELT Journal, 69*(3), 264–274. https://doi.org/10.1093/elt/ccv010

Lynch, B. K. (1996). *Language program evaluation.* Cambridge, MA: Cambridge University Press.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). California, CA: Sage.

Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art - and beyond. In S. Braun, J. Mukherjee, & K. Kohn (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods* (pp. 5–24). Frankfurt: Peter Lang Publishing Group.
Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
Özdemir, N. O. (2014). Using corpus data to teach collocations in medical English. *Journal of Second Language Teaching & Research, 3*(1), 37–52.
Rezaee, A. A., Marefat, H., & Saeedakhtar, A. (2014). The effect of data-driven learning (DDL) on Iranian EFL learners' acquisition of phrasal verbs. *Procedia - Social and Behavioral Sciences, 98*, 726–731. https://doi.org/10.1016/j.sbspro.2014.03.379
Schmidt, R. W. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp. 165–210). London, UK: Academic Press.
Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press
Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL, 26*(2), 184–201.
Tekin, B., & Soruç, A. (2016). Using corpus-assisted learning activities to assist vocabulary development in English. *TOJET: The Turkish Online Journal of Educational Technology, Special Issue*, 1270–1283.
Thurstun, J., & Candlin, C. N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes, 17*, 267–280.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.
Wu, S., Witten, I., & Franken, M. (2010). Utilizing lexical data from a Web-derived corpus to expand productive collocation knowledge. *ReCALL, 22*(1), 83–102.
Yılmaz, M., & Soruç, A. (2015). The effect of corpus-based activities on verb-noun collocations in EFL classes. *TOJET: The Turkish Online Journal of Educational Technology, 14*(2), 195–205.
Yoon, C. (2011). Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes, 10*(3), 130–139.
Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology, 12*(2), 31–48.

Copyrights