*ReCALL (2026), 38: 3, 324–344*
doi:10.1017/S0958344026100494

RESEARCH ARTICLE

# Tracing the diachronic effects of data-driven learning on lexical complexity in EFL learners' argumentative writing

Yanan Zhao
Shandong University, China (202220253@mail.sdu.edu.cn)

Jihua Dong
Shandong University, China (jdon104@aucklanduni.ac.nz; dongjihua@sdu.edu.cn)

**Abstract**

This study investigates the effectiveness of data-driven learning (DDL) in promoting lexical complexity in Chinese English as a foreign language (EFL) learners' argumentative writing, tracks developmental trajectories, and examines learners' perceptions. Adopting a quasi-experimental design, one class ($n = 26$) received DDL instruction, and the other ($n = 22$) received non-DDL instruction. Data were collected using triangulation, including argumentative writing samples from five time points, pre- and post-instruction questionnaires and semi-structured interviews. Results showed that learners in the DDL class significantly improved their lexical complexity, while the non-DDL class experienced declines. Across the five time points, nonlinear trajectories were observed in lexical complexity at the individual learner level. Learners reported positive attitudes toward DDL, though some challenges in corpus use remained. These findings provide empirical support for the effectiveness of DDL in promoting lexical complexity development in Chinese EFL learners' argumentative writing and provide pedagogical implications for corpus-based writing instruction.

**Keywords:** data-driven learning; lexical complexity; argumentative writing; longitudinal research

## 1. Introduction

Data-driven learning (DDL) is an approach that uses corpora in the classroom to promote learners' understanding of language patterns (Boulton & Vyatkina, 2024). DDL reflects constructivist principles (Cobb, 1999a) by engaging learners in exploring authentic data and constructing lexical knowledge through observation and inference (Boulton & Vyatkina, 2024). DDL also aligns with the noticing hypothesis (Schmidt, 1990), as concordance lines present lexical items and their context in a salient way (Liu & Gablasova, 2025). Pedagogical procedures for DDL commonly follow an inductive sequence from corpus observation to generalization, as reflected in Flowerdew's (2009) "four Is" of illustration, interaction, intervention, and induction, and Johns' (1991) identify-classify-generalize approach.

Lexical acquisition refers to the process of building large, well-structured, and richly interconnected lexical networks that encompass both knowing the meanings of words and

**Corresponding author:** Jihua Dong; Email: jdon104@aucklanduni.ac.nz

**Cite this article:** Zhao, Y. & Dong, J. (2026). Tracing the diachronic effects of data-driven learning on lexical complexity in EFL learners' argumentative writing. *ReCALL* 38(3): 324–344. https://doi.org/10.1017/S0958344026100494

© The Author(s), 2026. Published by Cambridge University Press on behalf of EUROCALL, the European Association for Computer-Assisted Language Learning. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

knowing the uses, collocations, and associations of words (Cobb, 1999b). In writing, lexical acquisition can be reflected in lexical complexity, which is defined as the degree of elaboration and diversity, and range of lexical items and collocations in a text, and is widely recognized as an effective indicator of L2 proficiency (Kyle et al., 2021). Lexical complexity comprises lexical sophistication and diversity, each reflecting a distinct aspect of lexical acquisition (Bulté & Housen, 2012). Lexical sophistication indicates learners' ability to use less frequent and later-acquired words (Kyle & Crossley, 2015). Lexical diversity reflects the breadth of productive vocabulary through the variety of distinct words used in a text (Nasseri & Thompson, 2021).

A plethora of previous DDL studies have examined and reported the positive effect of corpus use for language learning, particularly for lexical acquisition (e.g., Boulton & Cobb, 2017; Dong et al., 2023; Lee et al., 2019; Pérez-Paredes, 2022). In the domain of lexical complexity, previous studies have provided primary evidence for the positive impact of DDL on various lexical complexity measures such as frequency, variety, and type-token ratio (TTR; e.g., Muftah, 2023; Şahin Kızıl, 2023; Sun & Hu, 2023). These findings highlight the potential of DDL to promote learners' improvements in lexical complexity in authentic written production.

Despite growing interest in DDL, its effects on lexical complexity from a dynamic perspective remain underexplored. The term *dynamic* refers to the longitudinal and nonlinear nature of lexical development over time. Informed by complex dynamic systems theory (CDST), this study views L2 lexical development as a fluctuating process, where subsystems such as lexical diversity and sophistication fluctuate over time, and individuals may experience periods of rapid growth, plateaus, or regressions (Larsen-Freeman & Cameron, 2008). In this study, lexical complexity was operationalized in terms of lexical diversity and lexical sophistication. Lexical density was not included because it has been shown to relate more closely to register than to proficiency and is therefore rarely used as an indicator of lexical complexity (Díez-Ortega & Kyle, 2024; Engber, 1995; Linnarud, 1986).

## 2. Literature review

### 2.1. Research on the effectiveness of DDL for lexical acquisition

Previous studies have examined the effectiveness of DDL for vocabulary and collocation learning and have generally reported positive effects. These studies have assessed DDL effects through vocabulary tests (Chan & Liou, 2005; Karras, 2016), collocation tests (Liu & Gablasova, 2025; Saeedakhtar et al., 2020; Vyatkina, 2016), and comprehension and production tests (Frankenberg-Garcia, 2014). These findings suggested that DDL could affect not only learners' knowledge of individual words and collocations but also their productive use of lexical resources in writing. DDL may therefore help learners move beyond repetitive vocabulary and internalize more sophisticated lexical patterns, thereby fostering lexical diversity and sophistication in their writing. Examining DDL from the perspective of lexical complexity thus provides a theoretically grounded lens for understanding how corpus use shapes learners' productive vocabulary and pedagogical insights into how DDL can promote different aspects of lexical acquisition in L2 writing.

To our knowledge, only a small number of studies have utilized lexical complexity as the assessment of learners' lexical acquisition. For instance, Luo (2016) examined the effects of DDL activities on learners' writing performance using sophisticated TTR and found no significant improvements in complexity measures. Tsai (2021) examined the effects of DDL on EFL learners' lexical diversity in writing and reported learners' improvements in most measures, such as the number of different words and lexical word variation, with the exception of verb variation. Later, Samoudi and Modirkhamene (2022) compared direct DDL, indirect DDL, and non-DDL instruction across complexity, accuracy, and fluency dimensions, and reported that indirect DDL significantly improved accuracy and fluency, but not complexity. More recently, Hu and Deng

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

(2023) examined the impact of DDL writing practices on learners' lexical diversity via MTLD and lexical sophistication based on low-frequency word type ratio and found significant improvement in both indices. Şahin Kızıl (2023) incorporated lexical diversity measured by TTR and reported that DDL facilitated effective writing revision. Muftah (2023) used a similar measure and found comparable effects for BNCweb (Hoffmann et al., 2008) and Sketch Engine (Kilgarriff et al., 2014).

### 2.2. Research on the developmental pattern of lexical complexity

CDST has increasingly been adopted in second language acquisition (SLA) research as a lens to investigate linguistic complexity (Bulté & Housen, 2020). CDST conceptualizes language as a dynamic and adaptive system in which change is characterized by variability, nonlinearity, and sensitivity to contextual factors (Dong et al., 2023; Larsen-Freeman, 2009). Within this framework, language is composed of multiple interconnected subsystems including morphological, lexical, syntactic, and discourse features. Lexical complexity is often understood as developing through fluctuations, spurts, and regressions that reflect the interplay of multiple subsystems (Bulté & Housen, 2020). Importantly, CDST highlights variability itself as a central window into development, treating it not as statistical "noise" but as an integral part of the language learning process and a potential driver of growth (Larsen-Freeman, 2009).

A growing body of longitudinal research has examined developmental trajectories of lexical complexity across different learner groups and instructional contexts. These studies showed that developmental trajectories of lexical complexity were closely tied to instructional environments. More specifically, naturalistic exposure and immersion programs yielded fluctuations and regressions in the growth of lexical diversity (Pfenninger, 2020; Polat & Kim, 2014); classroom instruction showed various patterns of either linear or nonlinear development in different dimensions (Kim et al., 2025; Zheng, 2016); and study-abroad contexts contributed to learners' nonlinear growth of lexical diversity (Díez-Ortega & Kyle, 2024; McManus et al., 2021). For instance, Polat and Kim (2014) examined the development of lexical diversity among English learners of Turkish immigrants under naturalistic exposure and found nonlinear development in lexical diversity. McManus et al. (2021) similarly documented gradual but nonlinear growth in lexical diversity among advanced French and Spanish learners studying abroad. Zheng (2016) also reported nonlinear development in lexical sophistication and diversity among L2 learners of classroom instruction. Pfenninger (2020) also identified nonlinear growth in lexical diversity among Swiss learners in a bilingual immersion program, and Díez-Ortega and Kyle (2024) found steady linear increases in diversity but mixed, index-dependent trends in sophistication among UK students of Spanish studying abroad. By contrast, Kim et al. (2025) observed significant linear growth in sophistication but no change in diversity among beginning Japanese EFL students in a university English course.

Despite the growing body of research on lexical complexity development, limited attention has been paid to how DDL influences longitudinal development of lexical complexity. To address this, this study adopted a longitudinal, mixed-methods design involving two types of instruction: DDL and traditional instruction. Specifically, we examine (a) pretest–posttest comparisons under timed conditions, (b) holistic developmental trajectories across timed and untimed writing conditions, and (c) learners' perceptions of the DDL activities. The study is guided by the following research questions (RQs):

1. What is the effectiveness of DDL on the different dimensions of lexical complexity of learners' argumentative writing in timed conditions, as compared with traditional instruction?
2. What are the developmental trajectories of the different dimensions of lexical complexity of learners' argumentative writing in timed and untimed conditions under DDL instruction, as compared with traditional instruction?
3. How do learners perceive the effectiveness, experience, and challenges of DDL activities?

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

## 3. Method

### 3.1. Research setting and participants

The study involved undergraduate students in a compulsory English writing course in China, comprising a DDL class ($n = 26$) and a non-DDL class ($n = 22$). Based on their scores in the Test for English Majors Band 4 (TEM-4),[1] the participants' English proficiency was approximately aligned with the B1-B2 level of the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001). A pretest of argumentative writing was conducted before the intervention. An independent-samples *t*-test showed no significant difference between the DDL class ($M = 74.48$, $SD = 5.44$) and the non-DDL class ($M = 74.17$, $SD = 4.43$), $t(46) = 0.22$, $p = 0.826$, showing that the two classes were parallel in their writing proficiency. In the pre-instruction questionnaire, none reported prior use of a corpus for language learning.

### 3.2. Instructional procedures

The course spanned 18 weeks, with four hours of English instruction per week. Both classes followed the same curriculum and undertook the same assessments. Both the DDL and non-DDL classes were taught by the same instructor (one of the authors). Two hours each week were dedicated to the pedagogical intervention described below, while the remaining instructional time focused on general argumentative writing instruction. DDL and non-DDL instruction were embedded in regular class time (see Appendix A; all appendices appear in the supplementary material).

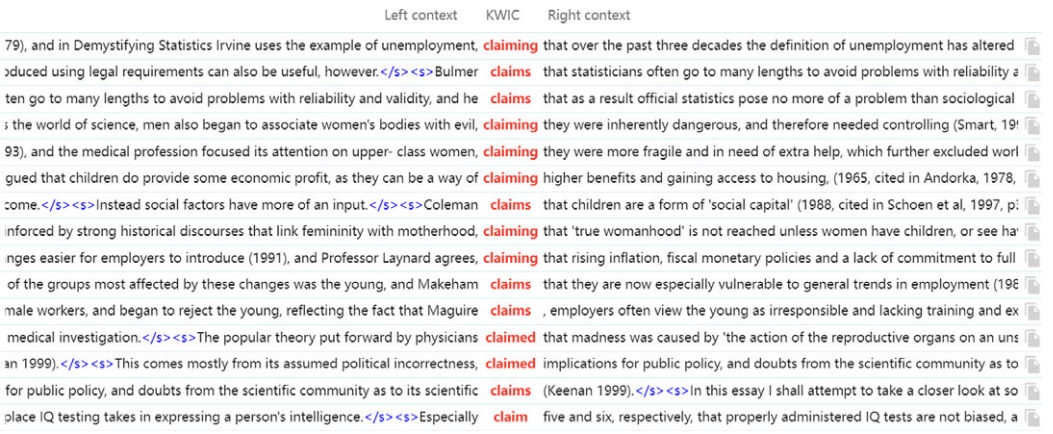
In the DDL instruction, learners used Sketch Engine, a corpus analysis platform chosen for its powerful functions such as wildcard searches and frequency comparisons (Crosthwaite & Steeples, 2024). At the beginning of the semester, a one-hour training session was provided to the DDL class to introduce the use of Sketch Engine, mainly focusing on the Concordance and WordSketch modules. The instructor modeled how to conduct searches and interpret the results, and students then practiced under guidance. During corpus-searching activities, learners primarily used the Concordance and WordSketch functions. Concordance (Figure 1) enabled observation of usage patterns in concordance lines, while WordSketch presented collocational profiles grouped by grammatical relations (Figure 2) with contexts (Figure 3).

Learners' corpus search was guided by worksheets, which were designed by the instructor based on previous DDL practices and pedagogical experience, and further refined through consultation with three other EFL teachers to ensure instructional clarity and appropriateness. The selection of lexical items adhered to the course syllabus and was informed by the prescribed textbook, *Writing Critically 3 for Argumentative Writing* (Chen et al., 2016), while also taking into account frequency information from the British Academic Written English Corpus (BAWE; Alsop & Nesi, 2009). For example, Units 2 and 3 of this textbook focus on making and supporting claims (see Appendix A); accordingly, target collocations for constructing claims and data (e.g., *assert a claim*, *provide evidence*) were incorporated into guided worksheets for classroom practice (see Appendix B for a sample worksheet). Students searched by themselves, followed by group discussions of two to four students, and were encouraged to share their findings with the whole class.

In the non-DDL class, students received instruction based on the same lexical items but followed the traditional presentation–practice–production (PPP) approach (Smart, 2014). In the presentation phase, the learners were introduced to the target vocabulary via slides. During the practice phase, they completed controlled exercises. In the production phase, they used the vocabulary in writing-related tasks such as drafting topic sentences.

[1]The TEM-4 is a standardized, criterion-referenced examination for undergraduate English majors in China.

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press



| Left context | KWIC | Right context |
|---|---|---|
| 79), and in Demystifying Statistics Irvine uses the example of unemployment, | claiming | that over the past three decades the definition of unemployment has altered |
| oduced using legal requirements can also be useful, however.</s><s>Bulmer | claims | that statisticians often go to many lengths to avoid problems with reliability a |
| ten go to many lengths to avoid problems with reliability and validity, and he | claims | that as a result official statistics pose no more of a problem than sociological |
| ; the world of science, men also began to associate women's bodies with evil, | claiming | they were inherently dangerous, and therefore needed controlling (Smart, 19 |
| 93), and the medical profession focused its attention on upper- class women, | claiming | they were more fragile and in need of extra help, which further excluded worl |
| gued that children do provide some economic profit, as they can be a way of | claiming | higher benefits and gaining access to housing, (1965, cited in Andorka, 1978, |
| come.</s><s>Instead social factors have more of an input.</s><s>Coleman | claims | that children are a form of 'social capital' (1988, cited in Schoen et al, 1997, p: |
| nforced by strong historical discourses that link femininity with motherhood, | claiming | that 'true womanhood' is not reached unless women have children, or see ha |
| nges easier for employers to introduce (1991), and Professor Laynard agrees, | claiming | that rising inflation, fiscal monetary policies and a lack of commitment to full |
| of the groups most affected by these changes was the young, and Makeham | claims | that they are now especially vulnerable to general trends in employment (198 |
| male workers, and began to reject the young, reflecting the fact that Maguire | claims | , employers often view the young as irresponsible and lacking training and ex |
| medical investigation.</s><s>The popular theory put forward by physicians | claimed | that madness was caused by 'the action of the reproductive organs on an uns |
| an 1999).</s><s>This comes mostly from its assumed political incorrectness, | claimed | implications for public policy, and doubts from the scientific community as to |
| for public policy, and doubts from the scientific community as to its scientific | claims | (Keenan 1999).</s><s>In this essay I shall attempt to take a closer look at so |
| place IQ testing takes in expressing a person's intelligence.</s><s>Especially | claim | five and six, respectively, that properly administered IQ tests are not biased, a |

**Figure 1.** The screenshot of the Sketch Engine interface showing a concordance search result "claim".

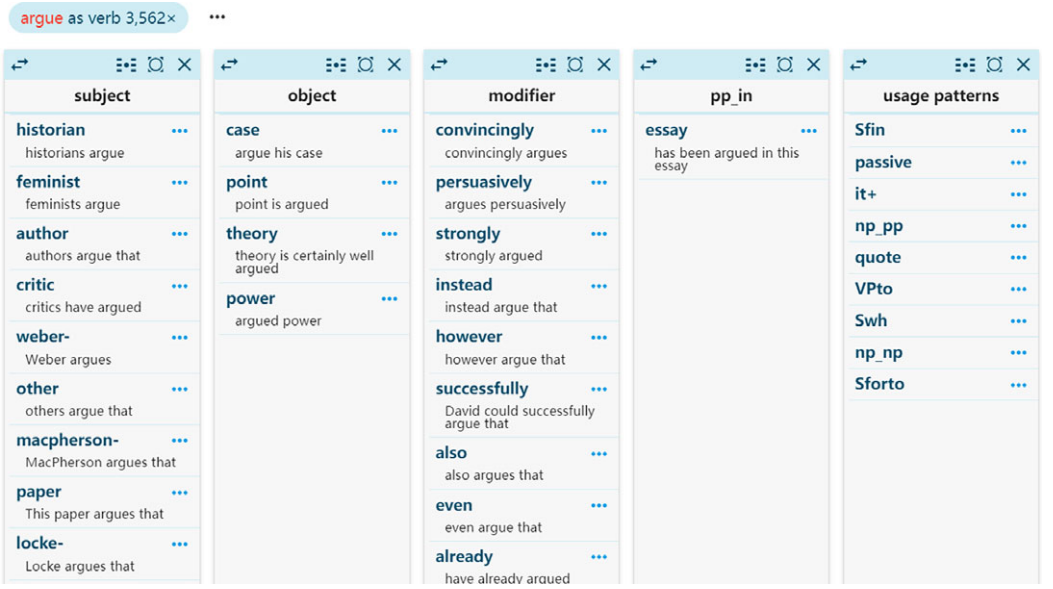



**Figure 2.** The screenshot of the Sketch Engine interface showing collocates of "argue".

### *3.3. Measures of lexical complexity*

Guided by the two dimensions of lexical complexity, we adopted four indices (see Table 1) as these indices have been attested to be effective in capturing different aspects of learners' lexical use and have been widely adopted in previous studies (e.g., Nasseri & Thompson, 2021; Zhou et al., 2023). TAALES (Kyle & Crossley, 2015) was used to measure lexical sophistication, and TAALED (Kyle et al., 2021) was used to assess lexical diversity. More explanations of these indices can be found in Appendix C.

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

**Table 1.** Indices for assessing the lexical complexity

| Dimension | Tool | Index | Explanation |
|---|---|---|---|
| Lexical sophistication | TAALES | *COCA_academic_frequency_log_aw* | COCA academic frequency (all words, log-transformed) |
| | | *Kuperman_AoA_aw* | Kuperman age of acquisition (all words) |
| Lexical diversity | TAALED | *mtld_original_aw* | Measure of textual lexical diversity (MTLD), original version (all words) |
| | | *mattr50_cw* | Moving-average type-token ratio (TTR), 50-word window (content words) |

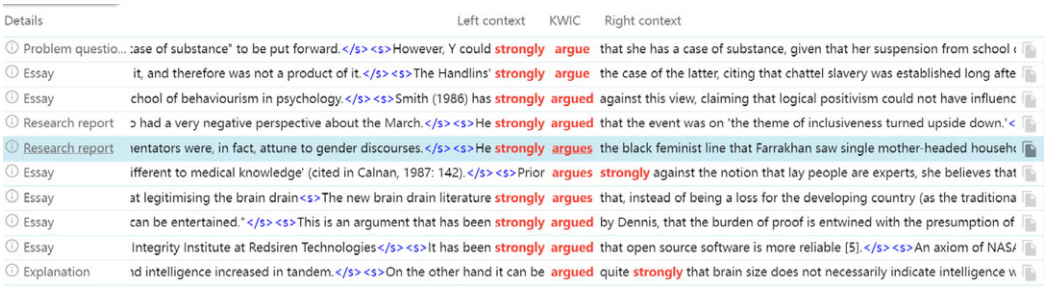



**Figure 3.** The screenshot of the Sketch Engine interface showing concordances of "strongly argued".

### *3.4. Data collection and analysis*

#### *3.4.1. Writing samples*

The diachronic effects of DDL on learners' lexical complexity were evaluated through writing samples. This study adopted both timed in-class writing (Week 1, 18) and untimed out-of-class essays (Week 5, 10, and 15) for a more comprehensive evaluation of students' writing abilities. The writing samples were collected before generative AI tools such as ChatGPT became publicly available; corpus tools were inaccessible during writing, and submissions were checked using plagiarism detection tools (more details can be found in Appendix D).

Before analysis, composite *z*-scores were computed to represent each learner's overall lexical complexity at each time point. Specifically, all four indices were *z*-standardized; the negatively oriented index (*COCA_academic_frequency_log_aw*) was reverse-coded, and the standardized indices were then averaged to form a composite score, such that higher composite *z*-scores indicated greater lexical complexity (see the detailed procedures in Appendix E). In comparing lexical complexity before and after instruction (RQ1), we used paired- and independent-samples *t*-tests to compare differences between the pretest and posttest and between the two classes at the pretest and the posttest, which were timed in-class writing tasks completed in Week 1 and 18. Effect size thresholds recommended by Plonsky and Oswald (2014) were adopted, namely 0.60 (small), 1.00 (medium), and 1.40 (large) for within-group designs and 0.40 (small), 0.70 (medium), and 1.00 (large) for between-group designs. When measuring the changes in lexical complexity over time (RQ2), a series of linear mixed-effects (LME) models were employed. For each lexical complexity index, we built three separate models to examine both within-group changes and between-group differences. These models were applied across five time points to capture developmental trajectories in the composite measure of overall lexical complexity and in the four individual indices representing lexical diversity and lexical sophistication throughout the intervention. The first two models were run separately for the DDL and non-DDL classes. Each model included time as a fixed factor, participant as a random intercept, random slopes for time,

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

**Table 2.** Descriptive statistics for overall lexical complexity

| Group | T1 | T2 | T3 | T4 | T5 |
|---|---|---|---|---|---|
| DDL | −0.127 (0.345) | −0.011 (0.479) | 0.073 (0.582) | 0.101 (0.556) | 0.182 (0.402) |
| Non-DDL | 0.150 (0.379) | 0.013 (0.502) | −0.088 (0.400) | −0.122 (0.460) | −0.198 (0.419) |

**Table 3.** Results of paired-samples *t*-tests for overall lexical complexity

| Group | *t* | *p* | *d* |
|---|---|---|---|
| Non-DDL | −2.303 | 0.033* | 0.515 |
| DDL | 3.695* | 0.001** | 0.788 |

*Note*: *$p <$ .05, **$p <$ .01, ***$p <$ .001.

**Table 4.** Results of independent-samples *t*-tests for overall lexical complexity

| Time | *t* | *p* | *d* |
|---|---|---|---|
| Pretest | 2.623 | 0.012* | 0.766 |
| Posttest | −3.133 | 0.003** | 0.926 |

*Note*: *$p <$ .05, **$p <$ .01, ***$p <$ .001.

and one lexical complexity index as the dependent variable. The third model included data from both classes, with time, group, and their interaction as fixed effects, and participant as a random intercept with random slopes for time. The three models allowed for evaluating trends in each pedagogical condition and group differences simultaneously. All models were built using the lme4 package (Bates et al., 2015) in R (R Core Team, 2021).

#### 3.4.2. Questionnaires

The pre-instruction questionnaire, administered in Week 1, collected learners' demographic information and prior corpus knowledge. The post-instruction questionnaire in Week 18 assessed learners' perceptions of DDL activities, following previous DDL research (Crosthwaite & Steeples, 2024; Yoon & Hirvela, 2004). The 14-item questionnaire covered perceived effectiveness, experience of corpus use, and ease and challenges of corpus use. Responses were recorded on a 5-point Likert scale and analyzed using percentages, means, and standard deviations.

#### 3.4.3. Interviews and reflective journals

To triangulate the quantitative findings, we collected monthly reflective journals from DDL participants and conducted semi-structured interviews with six randomly selected students (Appendix F). The qualitative data were analyzed thematically following Braun and Clarke (2006). The first and second authors coded the data independently, refined the codes into broader themes, and resolved discrepancies through discussion. Percentage agreement reached 89%, indicating strong interrater reliability (Rau & Shih, 2021).

## 4. Results

### 4.1. Learners' performances on overall lexical complexity

#### 4.1.1. The pre–post comparisons of overall lexical complexity

Table 2 presents descriptive statistics for overall lexical complexity (*z*-scores) across the five time points, and Tables 3 and 4 report the paired- and independent-samples *t*-test results for

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

**Table 5.** Results of within-group LME models with fixed factor of time for overall lexical complexity

| Class | Estimate (β) | *SE* | *t* | *p* |
|---|---|---|---|---|
| Non-DDL | −0.083 | 0.026 | −3.227 | 0.002** |
| DDL | 0.074 | 0.027 | 2.761 | 0.007** |

*Note*: *$p < .05$, **$p < .01$, ***$p < .001$.

**Table 6.** Results of between-group LME models for overall lexical complexity

| Effect | β | *SE* | *t* | *p* |
|---|---|---|---|---|
| Group | −0.351 | 0.131 | −2.678 | 0.008** |
| Time × Group | 0.155 | 0.040 | 3.898 | < 0.001*** |

*Note*: *$p < .05$, **$p < .01$, ***$p < .001$.

**Table 7.** Descriptive statistics for lexical sophistication

| Variable | Group | T1 | T2 | T3 | T4 | T5 |
|---|---|---|---|---|---|---|
| *COCA_academic_frequency_log_aw* | | | | | | |
| | DDL | 0.297 (0.980) | −0.075 (1.005) | −0.081 (1.093) | −0.188 (0.942) | −0.294 (0.954) |
| | Non-DDL | −0.351 (0.926) | 0.091 (1.007) | 0.097 (0.888) | 0.227 (1.041) | 0.321 (0.969) |
| *Kuperman_AoA_aw* | | | | | | |
| | DDL | −0.292 (1.005) | −0.101 (0.957) | −0.158 (1.084) | 0.258 (1.026) | 0.176 (1.187) |
| | Non-DDL | 0.345 (0.898) | 0.123 (1.057) | 0.191 (0.872) | −0.312 (0.891) | −0.191 (0.726) |

T1 (pretest) and T5 (posttest). The within-group analysis showed that the DDL class improved from T1 ($M = -0.127$, $SD = 0.345$) to T5 ($M = 0.182$, $SD = 0.402$), $t(25) = 3.695$, $p = 0.001$, $d = 0.788$, while the non-DDL class declined from T1 ($M = 0.150$, $SD = 0.379$) to T5 ($M = -0.198$, $SD = 0.419$), $t(21) = -2.303$, $p = 0.033$, $d = 0.515$. The between-group analysis also showed a significant difference at posttest, $t(46) = -3.133$, $p = 0.003$, $d = 0.926$, indicating that the DDL class achieved higher overall lexical complexity than the non-DDL class at posttest.

#### 4.1.2. Developmental trajectories of overall lexical complexity

The developmental trajectories of overall lexical complexity in the two classes are shown in Figure 4, where the red line represents the DDL class and the blue line represents the non-DDL class. As can be seen, learners in the DDL class showed a gradual increase in overall lexical complexity; in contrast, the non-DDL class exhibited a consistent decline. The within-group LME models (Table 5) illustrated that the change over time was significant for both classes (DDL: $\beta = 0.074$, $SE = 0.027$, $t = 2.761$, $p = 0.007$; non-DDL: $\beta = -0.083$, $SE = 0.026$, $t = -3.227$, $p = 0.002$). The between-group LME models (Table 6) showed a significant interaction between time and group, $\beta = 0.155$, $SE = 0.040$, $t = 3.899$, $p < .001$. Notably, from T3 onward, the DDL class maintained higher mean values than the non-DDL class, and the gap widened over time. This pattern suggests that overall lexical complexity developed more positively in the DDL class than in the non-DDL class.

As regards the developmental trajectories, at the group level, the mean *z*-scores showed a relatively linear trend across time, likely because it summarized multiple standardized indices. At the individual level, however, learners followed nonlinear trajectories, indicating that lexical complexity development was not uniformly linear across learners.

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

**Table 8.** Results of paired-samples *t*-tests for lexical sophistication

| Variable | Group | *t* | *p* | *d* |
|---|---|---|---|---|
| *COCA_academic_frequency_log_aw* | Non-DDL | −1.541 | 0.140 | 0.345 |
| | DDL | 2.243 | 0.036* | 0.478 |
| *Kuperman_AoA_aw* | Non-DDL | 1.646 | 0.116 | 0.368 |
| | DDL | −1.284 | 0.213 | 0.274 |

*Note*: *$p < .05$, **$p < .01$, ***$p < .001$.

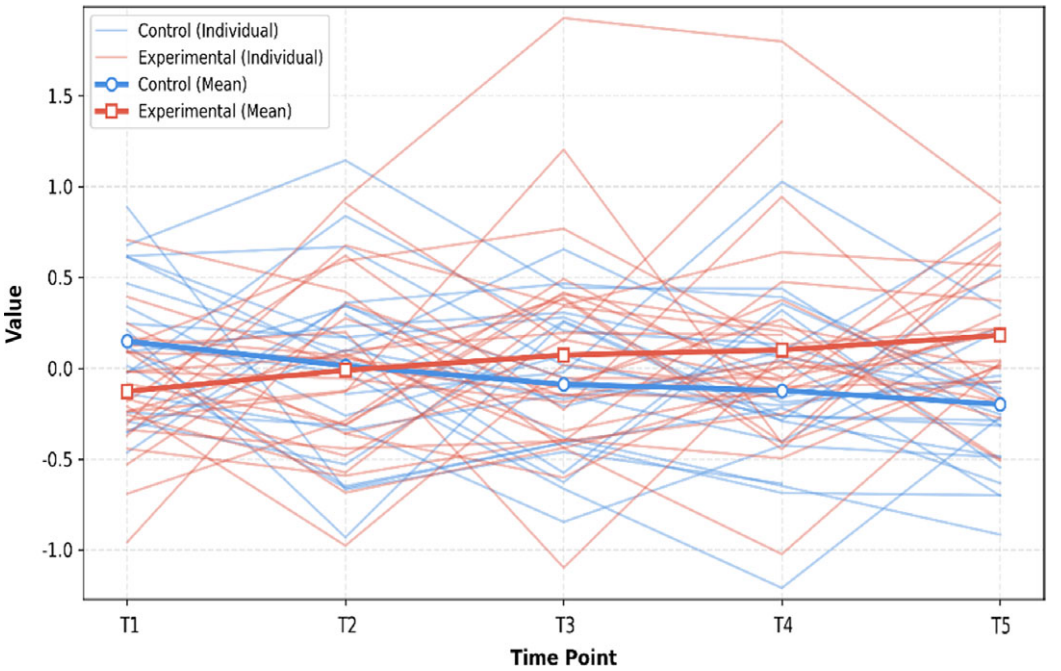



**Figure 4.** Developmental trajectories of the composite *z*-scores.

## 4.2. Learners' performances on lexical sophistication

### 4.2.1. The pre–post comparisons of lexical sophistication

Learners in the DDL class showed a significant increase in lexical sophistication, as evidenced by the significant decrease in the negatively oriented measure *COCA_academic_frequency_log_aw* from the pretest (T1: $M = 0.297$, $SD = 0.980$) to the posttest (T5: $M = -0.294$, $SD = 0.954$), $t(25) = 2.243$, $p = 0.036$, $d = 0.478$ (see Tables 7 and 8). In contrast, the non-DDL class showed a decrease in lexical sophistication, with the raw value increasing from T1 ($M = -0.351$, $SD = 0.926$) to T5 ($M = 0.321$, $SD = 0.969$), although this change was not statistically significant ($p = 0.140$). The between-group comparison (Table 9) further revealed a significant difference at posttest, $t(46) = 2.167$, $p = 0.036$, $d = 0.640$, showing the higher sophistication in the DDL class. For *Kuperman_AoA_aw*, no significant within-group changes were found in either class ($p > .05$). As shown by descriptive results in Table 7, the DDL class increased from T1 ($M = -0.292$, $SD = 1.005$) to T5 ($M = 0.176$, $SD = 1.187$), but the non-DDL class decreased from T1 ($M = 0.345$, $SD = 0.898$) to T5 ($M = -0.191$, $SD = 0.726$). These results suggest that the DDL class showed more positive development in lexical sophistication than the non-DDL class. Because corpus tools

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

**Table 9.** Results of independent-samples *t*-tests for lexical sophistication

| Variable | Time | *t* | *p* | *d* |
|---|---|---|---|---|
| *COCA_academic_frequency_log_aw* | Pretest | −2.353 | 0.023* | 0.678 |
| | Posttest | 2.167 | 0.036* | 0.640 |
| *Kuperman_AoA_aw* | Pretest | 2.316 | 0.025* | 0.665 |
| | Posttest | −1.277 | 0.209 | 0.369 |

*Note*: *$p < .05$, **$p < .01$, ***$p < .001$.

**Table 10.** Learners' writing examples from the posttest writing task

| No. | Class | Example |
|---|---|---|
| 1 | DDL | *From my perspective, policymakers should reconsider urban transport strategies, implement stricter regulations on private cars, and promote alternatives that could alleviate congestion.* |
| 2 | | *Setting up tidal lanes and changing lanes direction according to rush hour traffic is a good solution to alleviating congestion in high-density areas.* |
| 3 | | *So, from where I stand, privately owned cars should be banned in city centers as a practical solution, no matter how unrealistic and stringent some people may believe it is.* |
| 4 | Non-DDL | *We argue that the city should ban private cars and encourage people to take buses or trains instead, which can help alleviate the traffic congestion.* |
| 5 | | *More innovative road design can help direct the overflow of drivers onto side roads, reducing the chance of traffic congestion.* |
| 6 | | *It can not only ease traffic, but also reduce air and noise pollution. So banning private cars in the city center would definitely be a good decision.* |

were not available during the tests, this may indicate that learners successfully internalized and applied the lexical knowledge into independent writing.

Examples 1 and 2 (see Table 10) illustrate sentences produced by learners in the DDL class. These two examples demonstrate advanced lexical choices such as *reconsider*, *implement*, *regulations*, and *alleviate*, along with the phrase *from my perspective*, which reflect their high lexical sophistication. In contrast, Examples 4 and 5 come from students in the non-DDL class, which reflect more general lexical choices such as *ban*, *encourage*, *help*, and *take*, indicating a comparatively lower level of lexical sophistication.

### 4.2.2. Developmental trajectories of lexical sophistication

Figure 5 illustrates the developmental trajectories of lexical sophistication across five time points. The negative-oriented measure *COCA_academic_frequency_log_aw* showed a significant decrease over time in the DDL class, $\beta = -0.128$, $SE = 0.061$, $t = -2.086$, $p = 0.039$ (see Table 11), indicating increased use of less frequent vocabulary. In contrast, the non-DDL class showed a significant increase over time, $\beta = 0.147$, $SE = 0.064$, $t = 2.283$, $p = 0.024$. The interaction effect between time and group was also significant, $\beta = -0.275$, $SE = 0.089$, $t = -3.082$, $p = 0.002$ (see Table 12), indicating divergent trajectories between the two classes.

For *Kuperman_AoA_aw*, the DDL class exhibited a significant increase ($\beta = 0.125$, $SE = 0.059$, $t = 2.142$, $p = 0.034$), suggesting their use of more later-acquired words. There was a significant interaction effect ($\beta = 0.277$, $SE = 0.086$, $t = 3.203$, $p = 0.002$). As can be seen in Figure 5, the non-DDL class showed a decreasing pattern, while the DDL class increased over time. These results

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

**Table 11.** Results of within-group LME models with fixed factor of time for lexical sophistication

| Class | Variable | β | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| Non-DDL | *COCA_academic_frequency_log_aw* | 0.147 | 0.064 | 2.283 | .024* |
| | *Kuperman_AoA_aw* | −0.151 | 0.059 | −2.560 | .012* |
| DDL | *COCA_academic_frequency_log_aw* | −0.128 | 0.061 | −2.086 | .039* |
| | *Kuperman_AoA_aw* | 0.125 | 0.059 | 2.142 | .034* |

*Note*: *$p <$ .05, **$p <$ .01, ***$p <$ .001.

**Table 12.** Results of between-group LME models for lexical sophistication

| Variable | Effect | β | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| *COCA_academic_frequency_log_aw* | Group | 0.673 | 0.294 | 2.292 | .023* |
| | Time × Group | −0.275 | 0.089 | −3.082 | .002** |
| *Kuperman_AoA_aw* | Group | −0.852 | 0.284 | −2.997 | .003** |
| | Time × Group | 0.277 | 0.086 | 3.203 | .002** |

*Note*: *$p <$ .05, **$p <$ .01, ***$p <$ .001.

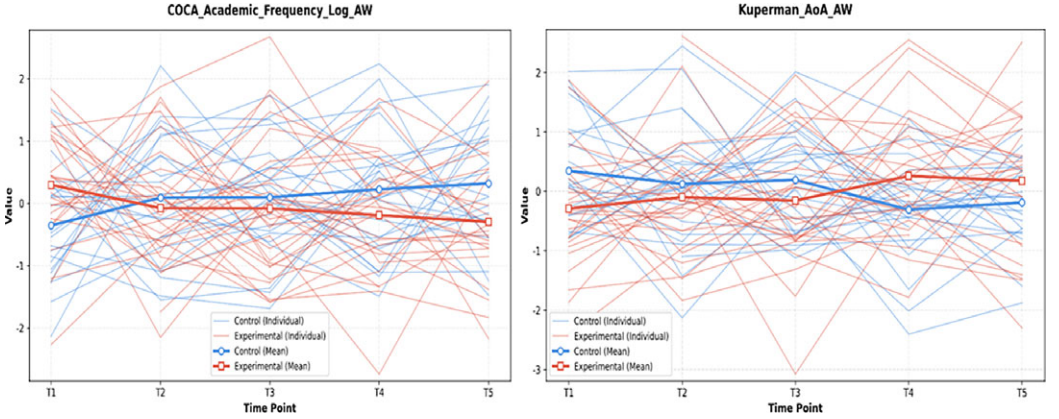



**Figure 5.** Developmental trajectories of lexical sophistication measures.

indicate that learners in the DDL class increasingly used more complex words over time, whereas those in the non-DDL class showed the opposite tendency.

The mean trajectories of *COCA_academic_frequency_log_aw* showed a continuous decrease in the DDL class, with a steeper decline from T1 to T2 followed by a more gradual decrease, whereas the non-DDL class showed a relatively rapid increase from T1 to T2 and a near-linear, gradual increase from T2 to T5. The mean trajectories of *Kuperman_AoA_aw* showed a small increase from T1 to T2 in the DDL class, slightly decreased from T2 to T3, increased from T3 to T4, and then decreased slightly from T4 to T5, whereas the non-DDL class decreased from T1 to T2, slightly increased from T2 to T3, decreased from T3 to T4, and then increased slightly from T4 to T5. At the individual learner level, trajectories showed fluctuation and nonlinearity over time. This suggests that lexical sophistication development was not uniformly linear across learners.

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

**Table 13.** Descriptive statistics for lexical diversity

| Variable | Group | T1 | T2 | T3 | T4 | T5 |
|---|---|---|---|---|---|---|
| *mtld_original_aw* | DDL | −0.026 (1.212) | −0.014 (0.974) | 0.275 (1.146) | 0.162 (1.014) | 0.309 (1.121) |
| | Non-DDL | 0.031 (0.698) | 0.017 (1.051) | −0.333 (0.672) | −0.195 (0.968) | −0.337 (0.734) |
| *mattr50_cw* | DDL | −0.203 (1.067) | 0.025 (1.077) | 0.259 (1.080) | 0.196 (0.886) | 0.267 (1.000) |
| | Non-DDL | 0.240 (0.879) | −0.030 (0.920) | −0.313 (0.808) | −0.236 (1.095) | −0.291 (0.937) |

**Table 14.** Results of paired-samples *t*-tests for lexical diversity

| Variable | Group | *t* | *p* | *d* |
|---|---|---|---|---|
| *mtld_original_aw* | Non-DDL | 1.791 | 0.089 | 0.401 |
| | DDL | −1.880 | 0.074 | 0.401 |
| *mattr50_cw* | Non-DDL | 1.738 | 0.099 | 0.279 |
| | DDL | −1.857 | 0.077 | 0.397 |

### 4.3. Learners' performances on lexical diversity

#### 4.3.1. The pre–post comparisons of lexical diversity

For lexical diversity (Table 13), the descriptive statistics showed an upward trend in *mattr50_cw* for the DDL class (T1: $M = -0.203$, $SD = 1.067$; T5: $M = 0.267$, $SD = 1.000$), whereas the non-DDL class showed a slight decline (T1: $M = 0.240$, $SD = 0.879$; T5: $M = -0.291$, $SD = 0.937$). Within each group, no statistically significant changes were found ($p > .05$; see Table 14), and the between-group comparison at posttest (see Table 15) was also not significant, $t(46) = -1.952$, $p = 0.057$, $d = 0.574$. Overall, the descriptive pattern suggests that the DDL class tended to show greater lexical diversity at posttest than the non-DDL class. This difference might be due to their sustained two-hour weekly engagement with corpus activities, which repeatedly exposed them to varied lexical patterns. Additional results for *mtld_original_aw* are provided in Appendix G in the supplementary material.

As exemplified in Example 3, lexical diversity was reflected in a varied range of vocabulary, including both general and academic words (e.g., *practical solution*, *unrealistic*, *stringent*). Example 6 from the non-DDL class illustrates more limited lexical diversity, with lexical items mostly being high-frequency and general words (e.g., *ease traffic*, *reduce*, *air and noise pollution*, *good decision*), with less use of precise, abstract, or domain-specific vocabulary.

#### 4.3.2. Developmental trajectory of lexical diversity

Table 16 shows the within-group LME results for lexical diversity, and Table 17 displays the between-group LME results. For lexical diversity, Figure 6 showed that the DDL class exhibited a gradual upward trajectory in *mattr50_cw*, whereas the non-DDL class showed a downward or more stable pattern over time. The between-group LME models showed significant interactions between time and group for *mattr50_cw* ($\beta = 0.237$, $SE = 0.085$, $t = 2.797$, $p = 0.006$), indicating that learners in the DDL class developed lexical diversity more positively than the non-DDL class.

The trajectories of mean values for *mattr50_cw* increased in an approximately linear pattern from T1 to T3 and then remained relatively stable from T3 to T5 in the DDL class, whereas in the non-DDL class, the mean score decreased from T1 to T3 and then remained relatively stable from

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

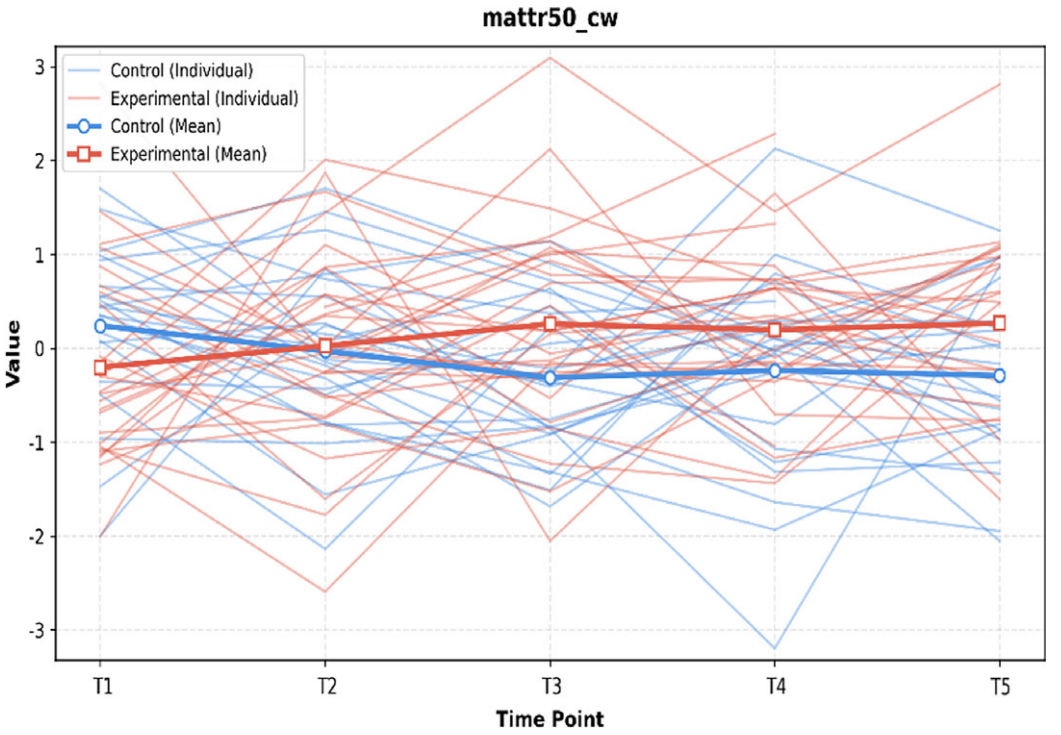



**Figure 6.** Developmental trajectories of *mattr50_cw*.

T3 to T5. At the individual level, learners followed highly varied and nonlinear paths, reflecting the dynamic and learner-specific nature of lexical diversity development over time, similar to the results of lexical sophistication.

### 4.4. Learners' perceptions of DDL activities

#### 4.4.1. Effectiveness of using corpora for argumentative writing

Figure 7 provides an overview of the three themes identified, namely the effectiveness of using corpora for argumentative writing, experience of corpus use, and easiness of using corpora. The results show that 82.61% ($M = 3.74$, $SD = 0.83$) of learners found corpora beneficial for improving collocation use, and 73.92% ($M = 3.70$, $SD = 0.82$) reported enhanced word choice in argumentative writing. Interview results further supported the benefits of DDL for word choice and accuracy, as illustrated in Excerpts 1 and 2 (Table 18). More than half of the learners (56.53%; $M = 3.57$, $SD = 0.79$) agreed that using corpora was beneficial for their argumentative writing. Specifically, one student, in the interview, explained that consulting corpora helped them choose lexical items that made their essay appear more coherent and well organized, thus improving their argumentative writing quality (see Excerpt 3).

#### 4.4.2. The use of corpora

The analysis showed that 60.87% of learners ($M = 3.43$, $SD = 0.98$) used corpora during writing practice and 78.26% ($M = 3.74$, $SD = 0.83$) used them to resolve writing difficulties, as illustrated in Excerpt 4. The results also showed that 65.21% ($M = 3.74$, $SD = 0.90$) could search corpora independently, 73.91% ($M = 3.57$, $SD = 1.00$) were confident in learning language use from corpora, and 73.91% ($M = 3.83$, $SD = 0.87$) evaluated the structured guidance positively (see

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

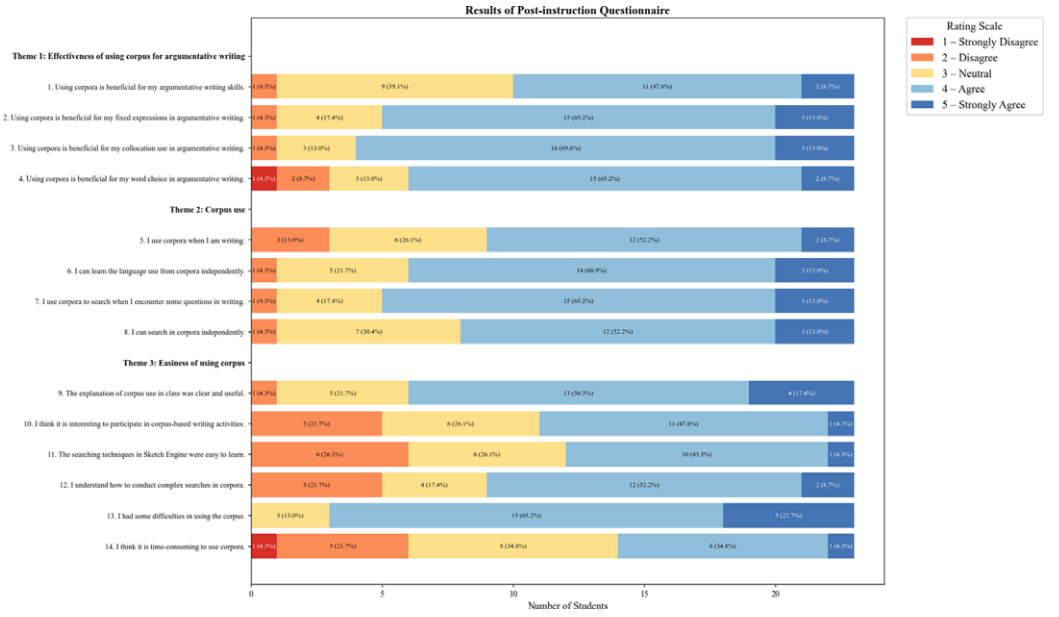



**Figure 7.** Results of post-instruction questionnaire.

Excerpts 5 and 6). In addition, 52.18% ($M = 3.57$, $SD = 1.00$) reported that corpus-based writing activities were interesting.

Despite the overall positive evaluations, several learners reported challenges in using corpus tools, particularly at the initial stages. The questionnaire results showed that 86.90% of learners encountered difficulties in using corpora, and 39.10% considered corpus consultation time-consuming.

## 5. Discussion

### 5.1. Effectiveness of DDL on lexical complexity

The analysis of the effectiveness of DDL on lexical complexity (RQ1) showed that the DDL class improved significantly in overall lexical complexity, whereas the non-DDL class showed a significant decline. More specifically, the results suggest that learners in the DDL class exhibited increased lexical sophistication, reflected in a greater use of less frequent vocabulary and words that are typically acquired at later stages of language development. This may be explained by their frequent engagement with concordance lines and authentic language input through corpus tools. This exposure may have helped learners internalize and use academic and low-frequency vocabulary. These findings are consistent with previous studies suggesting that DDL fostered the development of lexical complexity and supported the acquisition of academic and abstract vocabulary (Boulton & Cobb, 2017; Hu & Deng, 2023; Tsai, 2021). The regression observed in the non-DDL class may be attributed to the limited input provided in the PPP approach, which tends to lead learners to recycle familiar high-frequency vocabulary rather than extend their lexical repertoire (Harris & Leeming, 2022). The results further indicate that the DDL class demonstrated a stronger development in lexical diversity over time, particularly as reflected in the measures *mtld_original_aw* and *mattr50_cw*. This suggests that engagement with DDL activities may have supported learners in expanding the range of lexical items used in their writing across the instructional period. This benefit might be attributed to learners' exposure to a wider range of vocabulary resources, helping them avoid relying on repetitive vocabulary (Boulton & Cobb,

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

**Table 15.** Results of independent-samples *t*-tests for lexical diversity

| Variable | Time | *t* | *p* | *d* |
|---|---|---|---|---|
| *mattr50_cw* | Pretest | 1.577 | 0.122 | 0.449 |
| | Posttest | −1.952 | 0.057 | 0.574 |
| *mtld_original_aw* | Pretest | 0.203 | 0.84 | 0.056 |
| | Posttest | −2.331 | 0.025* | 0.676 |

*Note*: *$p < .05$, **$p < .01$, ***$p < .001$.

**Table 16.** Results of within-group LME models with fixed factor of time for lexical diversity

| Class | Variable | β | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| Non-DDL | *mattr50_cw* | −0.123 | 0.055 | −2.245 | .027* |
| | *mtld_original_aw* | −0.095 | 0.053 | −1.782 | 0.078 |
| DDL | *mattr50_cw* | 0.113 | 0.057 | 1.977 | 0.051 |
| | *mtld_original_aw* | 0.092 | 0.06 | 1.520 | 0.131 |

*Note*: *$p < .05$, **$p < .01$, ***$p < .001$.

**Table 17.** Results of between-group LME models for lexical diversity

| Variable | Effect | β | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| *mattr50_cw* | Group | −0.444 | 0.279 | −1.590 | 0.113 |
| | Time × Group | 0.237 | 0.085 | 2.797 | .006** |
| *mtld_original_aw* | Group | −0.214 | 0.284 | −0.753 | 0.452 |
| | Time × Group | 0.181 | 0.086 | 2.096 | .037* |

*Note*: *$p < .05$, **$p < .01$, ***$p < .001$.

2017). This is consistent with Şahin Kızıl (2023) and Muftah (2023), who reported that DDL significantly enhanced lexical diversity in L2 writing.

This finding corroborates previous DDL research suggesting that DDL is more effective in enhancing the acquisition of lexical patterns and collocations than traditional instruction (Boulton & Cobb, 2017; Frankenberg-Garcia, 2014; Lee et al., 2019; Liu & Gablasova, 2025; Saeedakhtar et al., 2020; Sun & Hu, 2023; Vyatkina, 2016), which may be explained by the underlying mechanisms of DDL. First, the KWIC format provided repeated exposure to authentic language data, helping learners notice and apply recurring lexical patterns (Frankenberg-Garcia, 2014; Pérez-Paredes, 2010). Second, the discovery-based nature of DDL positioned learners as active agents engaging with language data, which may enhance autonomy, motivation, and metalinguistic reflection (Johns, 1991). Third, repeated engagement with authentic corpus data may have facilitated incidental acquisition of academic and low-frequency vocabulary (Muftah, 2023).

### 5.2. Developmental trajectory of lexical complexity

This study also examined the developmental trajectories of lexical complexity under DDL and traditional instruction across five time points (RQ2). The results revealed clearly divergent patterns between the two classes: overall lexical complexity, lexical sophistication, and lexical diversity increased over time in the DDL class, whereas measures of these dimensions declined in

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

**Table 18.** Learner excerpts by theme

| Theme | No. | Excerpt |
|---|---|---|
| Effectiveness of using corpora for argumentative writing | 1 | *"I enjoy looking up synonyms as it helps me understand their differences, and makes my English more idiomatic and accurate." (Participant No. 1)* |
| | 2 | *"The corpus helps me find more formal and commonly used collocations, which prevents Chinglish expressions and makes my writing closer to native English speakers' habits." (Participant No. 2)* |
| | 3 | *"I think the improvement in clarity and flow is very noticeable. Through the corpus, I can clearly see how words are used to connect ideas, which helps me better express arguments and organize my writing." (Participant No. 3)* |
| The use of corpora | 4 | *"When I'm unsure about some collocations, it's easy to look them up, and you can also see which prepositions frequently accompany these words in different text types." (Participant No. 6)* |
| | 5 | *"It was hard to use at first, but I learned how to search better and use the results to improve my writing." (Participant No. 4)* |
| Easiness of using corpora | 6 | *"After learning the step-by-step search techniques, I now feel more confident using them in my writing." (Participant No. 5)* |

the non-DDL class. At the individual learner level, the developmental trajectories were not uniformly linear; rather, they exhibited fluctuations and nonlinear patterns across the observation period.

First, the results showed that the two classes followed diverging developmental trajectories across overall lexical complexity as well as sophistication and diversity indices. Specifically, the overall lexical complexity *z*-scores showed a linear mean-level increasing trajectory in the DDL class and a linear decreasing trajectory in the non-DDL class. A similar pattern was also observed for all measures of lexical sophistication, including corpus-based and psycholinguistic measures. For instance, the increase in *Kuperman_AoA_aw* showed that learners gradually used more advanced words, which were more abstract, lower in frequency, and cognitively more demanding. An increasing trajectory was also seen for lexical diversity in the DDL class; however, the non-DDL class decreased. Taken together, the trajectories suggest that DDL was more effective than traditional instruction in supporting the development of overall lexical complexity, lexical sophistication, and lexical diversity.

Second, the two lexical sophistication indices showed different mean-level trajectories over time. More specifically, the frequency-based index *COCA_academic_frequency_log_aw* showed a relatively steady, unidirectional pattern, whereas the psycholinguistic index *Kuperman_AoA_aw* showed greater fluctuation and nonlinearity. This difference may be related to the way the two indices are calculated. Frequency-based indices are based on the overall frequency values of many lexical items in a text and therefore provide a broader frequency profile, so changes in a small set of words are less likely to alter the mean score markedly across time points. By contrast, age-of-acquisition indices are more item-dependent and more sensitive to the lexical items sampled in a given text (De Wilde, 2023). As a result, the inclusion or absence of a small number of later-acquired words in a writing sample may lead to greater variation in mean-level trajectories across tasks (Zhou et al., 2023). This result is in line with Kim (2021), who found different trajectories for different lexical measures.

Third, at the individual learner level, all indices showed fluctuation and nonlinearity, with trajectories displaying variability and non-monotonic patterns over time. This is in line with previous studies showing that lexical sophistication and lexical diversity did not develop in a continuous linear manner and that individual trajectories often diverged from group-level trends

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

(Díez-Ortega & Kyle, 2024; Zheng, 2016). This variable and nonlinear pattern is also consistent with a CDST view, which regards variability as an inherent characteristic of L2 development rather than as random noise (Díez-Ortega & Kyle, 2024; Spoelman & Verspoor, 2010).

### 5.3. Learners' perceptions of DDL activities

The analysis of learners' perceptions found overall positive attitudes toward DDL instruction (RQ3). These observations are consistent with previous research showing that learners generally perceived corpus use as beneficial for developing L2 writing skills and useful for their L2 learning (Boulton, 2010; Yoon & Hirvela, 2004). Moreover, many learners reported successful application of corpus skills during writing, indicating that learners were able to transfer corpus-based strategies from classroom instruction to real-world writing tasks. A majority of participants expressed confidence in their ability to independently search corpora and interpret corpus output, suggesting that DDL fostered both corpus skills and writing confidence (Liu & Gablasova, 2025; Yoon & Hirvela, 2004). Learners also viewed corpus-based writing activities as interesting, in line with Boulton (2010).

Despite these benefits, some challenges were identified, particularly at the initial stage of DDL implementation. Learners reported difficulties with corpus interface design and search syntax, which made it challenging to locate relevant examples or interpret the results effectively (Dong & Wang, 2025; Liu & Gablasova, 2025; Saeedakhtar et al., 2020). This points to a broader concern in implementing DDL; that is, although corpus tools are pedagogically valuable, they are often regarded as cognitively demanding and time-consuming, particularly when learners lack sufficient training (Boulton, 2010; Zhao et al., 2026). Several learners in this study also emphasized that effective corpus use requires prior knowledge of grammar and word classes, which may pose a barrier for lower-proficiency learners. These findings support previous arguments that DDL tasks required cognitive and linguistic resources, which may overwhelm learners in the absence of structured support (Boulton & Cobb, 2017).

From a pedagogical standpoint, these difficulties underscore the need for explicit corpus literacy training, particularly at the early stages of DDL implementation. Teachers may need to carefully consider the potential challenges of introducing DDL, as learners who are unfamiliar with corpus tools may find it overwhelming. For instance, instructors can start with teacher-prepared corpus samples before transitioning students to independent searches, serving as a more accessible entry point for beginners (Sun & Hu, 2023). Once learners become familiar with corpus use, they tend to find it both effective and intrinsically motivating, highlighting the value of investing in training and task design (Boulton & Cobb, 2017).

## 6. Conclusion

This study investigated the effectiveness of DDL in promoting different dimensions of lexical complexity, explored their developmental trajectories, and examined learners' perceptions of DDL instruction. A comparative analysis revealed statistically significant differences, with DDL proving more effective in enhancing lexical complexity. Longitudinal analyses further uncovered diverse developmental trajectories across multiple dimensions of lexical complexity, characterized by fluctuations, regressions, and periods of stability. Moreover, the analysis of learners' responses to questionnaires and interviews indicated that, despite reported difficulties in using corpus tools, students generally perceived DDL as beneficial.

This study makes several contributions to the field of DDL. Theoretically, this study applies a multidimensional framework of lexical complexity that integrates both corpus-based and psycholinguistic indices. This framework provides a more nuanced approach to analyzing L2 lexical development, contributing to the understanding of lexical complexity in SLA. Empirically, the study suggests benefits of DDL for promoting lexical complexity and provides insights into

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

lexical development. Furthermore, the study's findings support the view that L2 lexical development is nonlinear, individualized, and dynamic. Learner trajectories did not follow uniform, linear progressions; instead, they may be influenced by individual differences and task-specific factors in L2 lexical acquisition, which may be worth further exploration. Pedagogically, the findings indicate that DDL has the potential to enhance more sophisticated and varied vocabulary use. At the same time, given the challenges students faced in using corpus tools, there is a need for scaffolding, gradual training, and adaptation of corpus-based tasks to learners' proficiency levels.

It is also necessary to point out several limitations. First, the small sample size limited the generalizability of the findings, and future research may consider including more diverse participants to validate the findings in a larger population. Second, while this study attempted to balance topic familiarity with the need to avoid topic repetition across tasks, the effect of topic could not be fully controlled. Future studies may consider adopting a more controlled design to reduce topic effect. Third, unauthorized assistance during learners' out-of-class writing tasks could not be entirely ruled out. Future research may consider using online timed writing platforms or tracking revision histories to enhance writing integrity.

**Supplementary material.** To view supplementary material for this article, please visit https://doi.org/10.1017/S0958344026100494

**Data availability statement.** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Authorship contribution statement.** Yanan Zhao: Methodology; Formal analysis; Writing – original draft. Jihua Dong: Supervision; Investigation; Writing – review & editing.

**Funding disclosure statement.** This work was supported by the National Social Science Foundation of China (No. 21FYYB052) and the Educational Reform Project of Shandong Province (Grant No. Z2022151).

**Competing interests statement.** The authors declare no competing interests.

**Ethical statement.** Ethical approval for this study was granted by Shandong University. Participation was voluntary, and informed consent was obtained from all participants before the study began. Participants were informed that their writing samples, questionnaire responses, interviews, and journals would be anonymized and used only for research purposes. No financial or non-financial competing interests affected the design, conduct, or reporting of the study.

**GenAI use disclosure statement.** During manuscript preparation, the authors used ChatGPT (OpenAI, accessible at https://chatgpt.com/) with GPT-5.2 in December 2025 and GPT-5.3 in March 2026 for limited language polishing and proofreading of author-written text. The tool was not used to generate data, analyze data, interpret results, or produce references. No modification or custom training of the tool was undertaken. All AI-assisted outputs were carefully reviewed and revised by the authors, who accept full responsibility for the content of the article as submitted.

## References

Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora*, 4(1), 71–83. https://doi.org/10.3366/E1749503209000227

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Vyatkina, N. (2024). Expanding methodological approaches in DDL research. *TESOL Quarterly*, 58(3), 1193–1204. https://doi.org/10.1002/tesq.3269

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.32.02bul

Bulté, B., & Housen, A. (2020). A critical appraisal of the CDST approach to investigating linguistic complexity in L2 writing development. In G. G. Fogal & M. H. Verspoor (Eds.), *Complex dynamic systems theory and L2 writing development* (pp. 207–238). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.54.09bul

Chan, T., & Liou, H.-C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb–noun collocations. *Computer Assisted Language Learning*, 18(3), 231–251. https://doi.org/10.1080/09588220500185769

Chen, Y. P., Lü, Y. B., & Liu, L. (Eds.). (2016). *Writing critically 3 for argumentative writing*. Foreign Language Teaching and Research Press.

Cobb, T. (1999a). Applying constructivism: A test for the learner-as-scientist. *Educational Technology Research & Development*, 47(3), 15–31. https://doi.org/10.1007/BF02299631

Cobb, T. (1999b). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning*, 12(4), 345–360. https://doi.org/10.1076/call.12.4.345.5699

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press. https://www.coe.int/lang-CEFR

Crosthwaite, P., & Steeples, B. (2024). Data-driven learning with younger learners: Exploring corpus-assisted development of the passive voice for science writing with female secondary school students. *Computer Assisted Language Learning*, 37(5–6), 1166–1197. https://doi.org/10.1080/09588221.2022.2068615

De Wilde, V. (2023). Lexical characteristics of young L2 English learners' narrative writing at the start of formal instruction. *Journal of Second Language Writing*, 59, Article 100960. https://doi.org/10.1016/j.jslw.2022.100960

Díez-Ortega, M., & Kyle, K. (2024). Measuring the development of lexical richness of L2 Spanish: A longitudinal learner corpus study. *Studies in Second Language Acquisition*, 46(1), 169–199. https://doi.org/10.1017/S0272263123000384

Dong, J., Liu, Y., & Lu, X. (2023). A discourse dynamics analysis of academic voice construction: Disciplinary variation, trajectories, and dynamic interaction patterns. *System*, 119, 103181. https://doi.org/10.1016/j.system.2023.103181

Dong, J., & Wang, H. (2025). Exploring the effectiveness of a DDL-GenAI teaching mode in students' EFL writing classroom. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2025.2577369

Dong, J., Zhao, Y., & Buckingham, L. (2023). Charting the landscape of data-driven learning using a bibliometric analysis. *ReCALL*, 35(3), 339–355. https://doi.org/10.1017/S0958344022000222

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4(2), 139–155. https://doi.org/10.1016/1060-3743(95)90004-7

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, 14(3), 393–417. https://doi.org/10.1075/ijcl.14.3.05flo

Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL*, 26(2), 128–146. https://doi.org/10.1017/S0958344014000093

Harris, J., & Leeming, P. (2022). The impact of teaching approach on growth in L2 proficiency and self-efficacy: A longitudinal classroom-based study of TBLT and PPP. *Journal of Second Language Studies*, 5(1), 114–143. https://doi.org/10.1075/jsls.20014.har

Hoffmann, S., Evert, S., Smith, N., Lee, D., & Berglund Prytz, Y. (2008). *Corpus linguistics with* BNCweb – *A practical guide*. Peter Lang.

Hu, K., & Deng, Y. (2023). The effect of corpus-based writing practices on EFL learners' lexical diversity and lexical sophistication. In C. Li, S. K. S. Cheung, F. L. Wang, A. Lu, & L. F. Kwok (Eds.), *Blended learning: Lessons learned and ways forward: 16th International Conference on Blended Learning, ICBL 2023, Hong Kong, China, July 17–20, 2023, proceedings* (pp. 85–96). Springer. https://doi.org/10.1007/978-3-031-35731-2_9

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal*, 4, 1–16.

Karras, J. N. (2016). The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. *ReCALL*, 28(2), 166–186. https://doi.org/10.1017/S0958344015000154

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7–36. https://doi.org/10.1007/s40607-014-0009-9

Kim, M. (2021). Exploring longitudinal changes in lexical and syntactic features in beginning-level EFL learner writing. *System*, *103*, Article 102680. https://doi.org/10.1016/j.system.2021.102680

Kim, M., Kim, Y., & Kang, S. (2025). The dynamics of changes in linguistic complexity and writing scores in timed argumentative writing among beginning-level EFL learners. *International Review of Applied Linguistics in Language Teaching*, 63(4), 2891–2925. https://doi.org/10.1515/iral-2023-0287

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757–786. https://doi.org/10.1002/tesq.194

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity using direct judgements. *Language Assessment Quarterly*, 18(2), 154–170. https://doi.org/10.1080/15434303.2020.1844205

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 579–589. https://doi.org/10.1093/applin/amp043

Larsen-Freeman, D., & Cameron, L. (2008). Research methodology on language development from a complex systems perspective. *The Modern Language Journal*, 92(2), 200–213. https://doi.org/10.1111/j.1540-4781.2008.00714.x

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, 40(5), 721–753. https://doi.org/10.1093/applin/amy012

Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish learners' written English*. Gleerup.

Liu, T., & Gablasova, D. (2025). Data-driven learning of collocations by Chinese learners of English: A longitudinal perspective. *Computer Assisted Language Learning*, 38(3), 612–637. https://doi.org/10.1080/09588221.2023.2214605

Luo, Q. (2016). The effects of data-driven learning activities on EFL learners' writing development. *SpringerPlus*, 5(1), Article 1255. https://doi.org/10.1186/s40064-016-2935-5

McManus, K., Mitchell, R., & Tracy-Ventura, N. (2021). A longitudinal study of advanced learners' linguistic development before, during, and after study abroad. *Applied Linguistics*, 42(1), 136–163. https://doi.org/10.1093/applin/amaa003

Muftah, M. (2023). Data-driven learning (DDL) activities: Do they truly promote EFL students' writing skills development? *Education and Information Technologies*, 28(10), 13179–13205. https://doi.org/10.1007/s10639-023-11620-z

Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, *47*, Article 100511. https://doi.org/10.1016/j.asw.2020.100511

Pérez-Paredes, P. (2010). Corpus linguistics and language education in perspective: Appropriation and the possibilities scenario. In T. Harris & M. Moreno Jaén (Eds.), *Corpus linguistics in language teaching* (pp. 53–73). Peter Lang. https://doi.org/10.3726/978-3-0351-0166-9

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

Pfenninger, S. E. (2020). The dynamic multicausality of age of first bilingual language exposure: Evidence from a longitudinal content and language integrated learning study with dense time serial measurements. *The Modern Language Journal*, 104(3), 662–686. https://doi.org/10.1111/modl.12666

Plonsky, L., & Oswald, F. L. (2014). How big Is "big"? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. https://doi.org/10.1111/lang.12079

Polat, B., & Kim, Y. (2014). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics*, 35(2), 184–207. https://doi.org/10.1093/applin/amt013

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rau, G., & Shih, Y.-S. (2021). Evaluation of Cohen's kappa and other measures of inter-rater agreement for genre analysis and other nominal data. *Journal of English for Academic Purposes*, 53, Article 101026. https://doi.org/10.1016/j.jeap.2021.101026

Saeedakhtar, A., Bagerin, M., & Abdi, R. (2020). The effect of hands-on and hands-off data-driven learning on low-intermediate learners' verb-preposition collocations. *System*, *91*, Article 102268. https://doi.org/10.1016/j.system.2020.102268

Şahin Kızıl, A. (2023). Data-driven learning: English as a foreign language writing and complexity, accuracy and fluency measures. *Journal of Computer Assisted Learning*, 39(4), 1382–1395. https://doi.org/10.1111/jcal.12807

Samoudi, N., & Modirkhamene, S. (2022). Concordancing in writing pedagogy and CAF measures of writing. *International Review of Applied Linguistics in Language Teaching*, 60(3), 699–722. https://doi.org/10.1515/iral-2020-2014

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. https://doi.org/10.1093/applin/11.2.129

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

Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics*, 31(4), 532–553. https://doi.org/10.1093/applin/amq001

Sun, X., & Hu, G. (2023). Direct and indirect data-driven learning: An experimental study of hedging in an EFL writing class. *Language Teaching Research*, 27(3), 660–688. https://doi.org/10.1177/1362168820954459

Tsai, Y.-R. (2021). Exploring the effects of corpus-based business English writing instruction on EFL learners' writing proficiency and perception. *Journal of Computing in Higher Education*, 33(2), 475–498. https://doi.org/10.1007/s12528-021-09272-4

Vyatkina, N. (2016). Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL*, 28(2), 207–226. https://doi.org/10.1017/S0958344015000269

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257–283. https://doi.org/10.1016/j.jslw.2004.06.002

Zhao, Y., Buckingham, L., & Dong, J. (2026). Enhancing learners' genre competence in proposal writing with data-driven learning. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2026.2644505

Zheng, Y. (2016). The complex, dynamic development of L2 lexical use: A longitudinal study on Chinese learners of English. *System*, 56, 40–53. https://doi.org/10.1016/j.system.2015.11.007

Zhou, X., Gao, Y., & Lu, X. (2023). Lexical complexity changes in 100 years' academic writing: Evidence from *Nature* Biology Letters. *Journal of English for Academic Purposes*, *64*, Article 101262. https://doi.org/10.1016/j.jeap.2023.101262

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press

## About the authors

**Yanan Zhao** is a PhD student in the School of Foreign Languages and Literature at Shandong University, China. She holds a master's degree in languages and linguistics from the University of Melbourne, Australia. Her research interests include data-driven learning, second language teaching and learning, and corpus linguistics. She has published in *ReCALL*, *Computer Assisted Language Learning*, and *Assessing Writing*.

**Jihua Dong** is a Professor, Qilu Young Scholar, and Taishan Young Scholar at Shandong University, China. Her research interests include corpus linguistics, data-driven learning and teaching, and academic writing. She has published in *Applied Linguistics*, *Assessing Writing*, *English for Specific Purposes*, *Journal of English for Academic Purposes*, *ReCALL*, and *System, etc.*

https://doi.org/10.1017/S0958344026100494 Published online by Cambridge University Press