

Fluency Development in Timed Writing: The Role of Proficiency

May 2026 – Volume 30, Number 1
<https://doi.org/10.55593/ej.30117a7>

Timothy Doe

Meiji University, Japan
<timdoe@meiji.ac.jp>

Teruyo Nakao

Toyo Gakuen University, Japan
<teruyo.nakao@tyg.jp>

Abstract

While research examining foreign language fluency development in timed writing has shown its potential to bring about improvement in language proficiency, it remains unclear how different groups of learners can benefit from this activity. To investigate the role of proficiency, this study examined the writing produced by 35 Japanese English as a foreign language university students from low- and higher-level classes. 422 samples of timed writing were collected over 13 weeks and analyzed for development in complexity, accuracy, lexis, and fluency (CALF). Results showed significant improvement in writing fluency for both groups. However, the higher-level group made steady gains across the semester, while growth in the lower-level group was limited to the initial weeks of the study. Furthermore, examination of the developmental trajectories of the CALF measures suggested the higher-level group was able to attend to both fluency and complexity, but the lower-level group was unable to do so. Both groups showed a great deal of variation in lexis and accuracy. The results suggest that learners need to pass a proficiency threshold in order for timed writing to benefit fluency development. Finally, although both groups reacted positively to the activity, differences were also seen in their reasoning.

Keywords: Fluency, Writing, Timed writing, Accuracy, Vocabulary

To date, much of the research into second language (L2) fluency development, the ability to produce language smoothly and with relative ease, has focused on oral speech. Investigations of written output remain relatively rare and little is known about the ways that students can become more fluent writers in foreign language learning contexts (Nguyen, 2015). However, learning more about how learners can develop their abilities to produce written texts efficiently is an important area of L2 research because many language learners find writing a time-consuming and difficult experience. Fluency development activities in university contexts have been criticized for potentially confusing students about the standards of academic writing and being misrepresentative of authentic university written assignments (Hinkel, 2004). However, research has shown that practice with less demanding writing tasks can lead to improvements

in producing longer texts (Nguyen, 2015; Yasuda, 2025) and the use of more complex language (Nitta & Baba, 2014), which in the long term may help with the development of writing skills. Nguyen (2015) found that third-year university English as a foreign language (EFL) students made larger fluency gains than those in their first year, although the pre-post design of the study makes it unclear how fluency improved. This issue is important to resolve because research has shown that learners can achieve a stable ceiling in their writing fluency and develop other dimensions of their writing proficiency (Nitta & Baba, 2014). Furthermore, it is possible that fluency development activities, during which learners are encouraged to focus on meaning, might lead to the automatization of incorrect language forms, a concern raised in L2 speaking studies that showed learners produced less accurate language in time-pressured conditions compared to those in constant time conditions (Boers, 2014; Thai & Boers, 2016). However, perhaps due to the difficulty in analyzing large quantities of student writing, little is known about how accuracy levels may change when learners participate in written fluency development activities over longer periods of time.

The aim of this study is to examine the degree to which students of differing proficiency levels are able to develop their writing fluency in timed-writing activities over the course of an academic semester, while observing how other aspects of language performance change.

Literature Review

Writing Fluency

Fluency is a multi-dimensional construct and has been operationalized in numerous ways in L2 studies (Segalowitz, 2010). Researchers have shown that oral fluency measures can be separated into three factors: speed (the number of words or syllables produced in a specified amount of time), breakdown (measures of pausing behaviors), and repairs (repetitions, reformulations, and corrections) (Tavakoli, 2025). Research into writing fluency, however, has tended to reduce fluency to one measure, the total number of words produced in a specified amount of time (Min, 2023), which is a problematic representation of writing fluency, as it captures the product of writing rather than the process (Abdel Latif, 2013). Alternative operationalizations are available, such as pause-related *P-bursts* and revision-related *R-bursts* (Chenoweth & Hayes, 2001), and key-logging software that can record a writer's pausing behaviors while typing (Cho, 2017). In an interpretive case study, Min (2023) also found that a teacher conceptualized writing fluency as the ability to utilize the resources necessary for higher order thinking while writing. Yet while these more fine-grained measures can provide a closer analysis of writing fluency, text length remains popular due to the difficulties involved in obtaining alternative measures, such as video recorders for *P-bursts* and *R-bursts*, and computer equipment (along with adjustments for participants' typing speed) for key-logging software. Therefore, while word counts of text length provide a superficial view of the construct, it is a practical measure that is capable of identifying differences among students, and due to its popularity, allows for comparisons with other studies (Yasuda, 2025).

The majority of writing fluency research has focused on general language proficiency rather than specific written skills (such as text organization or critical thinking), which is likely due to the importance of output for L2 development. Swain's output hypothesis (Swain & Lapkin, 1995) noted several benefits of the provision of output activities for learners, such as 'pushed output', when learners must switch from semantic processing (that is often sufficient to gain comprehension), to syntactic processing, a necessary condition for accurate speech or writing. It has been argued that output activities can result in fluency development, and goal-directed practice activities of previously learned skills are key to improvement, as growth takes place through the gradual proceduralization of declarative knowledge of the L2. This builds automaticity, meaning that over time, learners can develop the ability to use their language

knowledge with more speed and less difficulty (De Bot, 1996; DeKeyser, 2007a). However, in addition to a sufficient quantity of practice, in order for fluency development to take place, care must be taken to provide activities that are within learners' range of abilities (DeKeyser, 2007b), and designing activities that are relatively easy for learners to use previously learned language items is emphasized in L2 pedagogical literature (Macalister & Nation, 2020).

While it can be challenging for teachers of foreign languages to provide learners with sufficient quantities of practice due to reduced opportunities for contact with the L2, task repetition (TR), that is, the provision of repeated opportunities to perform certain tasks under the same or similar conditions (Bygate, 2001) is one means of doing so, and several studies have shown that TR can have a significant effect on L2 development (Abdi Tabari & Golparvar, 2024; Bygate, 2018; Fukunaga, 2023; Kim et al., 2020). In a study comparing TR and task rehearsal conditions (Abdi Tabari & Golparvar, 2024), learners from three different proficiency levels who repeated a writing task after a one-week interval were able to make significant gains in complexity, accuracy, lexis, and fluency (CALF) measures regardless of the condition. However, across the conditions, differences were found among the proficiency levels, with the higher-level group significantly outperforming the low- and mid-level students in most of the CALF measures, including fluency. Such studies illustrate how TR can help learners access more of their linguistic knowledge yet also show that gains may be affected by proficiency levels. However, in addition to these studies, which often use a pre- and post-test design, longitudinal studies are also needed to learn more about how writing develops over time for learners of different proficiency levels when repeating a task.

Several longitudinal studies of L2 writing have been conducted with timed-writing activities (also sometimes referred to as freewriting or speed writing), where learners produce a written text in a set time while tracking their word count, aiming to gradually increase the number of words that they can produce. These activities have also been argued to be effective for fluency development (Nation, 2008). Both quasi-experimental and descriptive studies of TR and language proficiency have shown that timed writing can lead to significant improvement in writing fluency (Nguyen, 2015; Yasuda, 2025) and other aspects of language proficiency (Nitta & Baba, 2014). The implementation of timed-writing activities can differ greatly, which likely affects the degree to which learners focus on fluency. For example, some studies allow for dictionary use (Nitta & Baba, 2014) and include relatively sophisticated topics (e.g., historical periods) (Baba & Nitta, 2014), while others use more familiar topics (e.g., school life) and explicitly instruct students to write as much as possible without paying attention to grammatical correctness (Choi & Kim, 2025; Yasuda, 2025). Research has shown that learners who engage in timed-writing can outperform those spending time on other language-learning activities. Penn and Lim (2016) found that over an academic semester, Korean university students who completed weekly 10-minute timed-writing activities made significantly more gains in their English proficiency than a control group. Finally, while timed-writing activities appear to be the most predominant means of examining TR and writing development, researchers have also used other written formats, such as descriptive and argumentative essays (Fukunaga, 2023), and blog posts (Kim et al., 2020).

CALF Measurements of L2 Timed Writing

While fluency development is an important aspect of L2 writing to explore, it should be looked at alongside other measures of writing proficiency, specifically, accuracy, lexis, and complexity, because improvements in one aspect of language proficiency may occur with changes in others (Skehan, 2009). In second language acquisition (SLA) research, many studies have employed the CALF framework to obtain a broader view of language development and/or the effects of interventions (Housen et al., 2012). Complexity refers to the range and

sophistication of structure and syntax and while it can be operationalized in a multitude of ways, common measures are the degree of clausal subordination and the length of clauses or sentences (Wolfe-Quintero et al., 1998). Accuracy measures look at the degree to which a text conforms to grammatical correctness and is often examined by measures such as the number of errors per clause and the ratio of error-free clauses to those containing errors (Wolfe-Quintero et al., 1998). Lexis measures examine the range and sophistication of vocabulary displayed in a text (Skehan, 2009). Finally, while fluency has been shown to have separate dimensions related to speed, breakdowns, and repairs (Tavakoli & Skehan, 2005), due to the challenges associated with recording temporal aspects of language production in timed-writing activities, researchers have operationalized fluency as the number of words produced in a set amount of time (Nguyen, 2015; Nitta & Baba, 2014; Yasuda, 2025). Research has shown that the design of language-learning activities can be manipulated to facilitate higher levels of performance in one or several of the CALF components (Skehan & Foster, 2012), giving insight into how a balanced approach to language learning might be implemented. Finally, it should be emphasized that even multiple measures in all CALF dimensions do not capture every aspect of a text, and other qualitative features, such as functional adequacy (Pallotti, 2009) are needed to examine the quality of text content and organization.

Several studies have investigated the development of CALF measures in repeated timed-writing activities. Baba and Nitta (2010) and Nitta and Baba (2014) examined lower-level Japanese university EFL students' language in 10-minute freewriting tasks. These studies are notable because data were collected at 30 different interval points over an academic year and show the developmental trajectories of a variety of CALF measures (excluding accuracy). Baba and Nitta (2010) compared the first and last texts and found significant change in two measures: sentence length and sentence syntax similarity. A closer analysis of the same group in Nitta and Baba (2014) found that fluency (operationalized by text length) peaked in the seventh week of the year, although significant improvements were found for complexity and lexis. One reason for the lack of fluency development may be related to the study design that allowed students to use dictionaries while they wrote, a feature that could be viewed as a type of online planning, that is associated with lower levels of fluency and higher levels of complexity (Ellis, 2005). In addition, the study examined specific TR (i.e., using the exact same writing topic for 2 consecutive weeks), however, no significant changes were found when comparing these texts, suggesting that in contrast to oral fluency research, writing fluency improves through repetitions of the same *types* of tasks (i.e., writing about different topics), rather than the same specific tasks. In a related study, Hosoda (2018) compared TR conditions with a control group in a 7-minute timed-writing task over 10 weeks. Results showed significant gains in fluency for all groups, and in contrast to Nitta and Baba (2014), significant effects for TR with several topics.

Follow-up studies have revealed more insight into the effects of timed writing. Baba and Nitta (2021) used a general growth mixture model to identify the text length developmental trajectories of 105 students. The most parsimonious model identified three groups: stagnating ($n = 57$), steadily increasing ($n = 45$) and markedly increasing ($n = 3$). The study revealed that members of the 2 increasing text length groups also received higher course grades and were more engaged in the task. From these results, the authors argue that certain learners may benefit from task scaffolding to better develop their writing skills. A case study involving two learners whose fluency improved more than other classmates (Baba & Nitta, 2014), showed that text length developed in a non-linear pattern, and identified several phase transitions (a discontinuous change resulting from the L2 system being reorganized into a higher-level state) that led to increases in text length. These transitions differed for the two learners, suggesting that writing development follows individual patterns. Consistent with Baba and Nitta (2021),

the results also showed that students were engaged with the writing task and reflected deeply on their performance.

Other studies have also shown that timed-writing activities can facilitate writing fluency development in EFL settings. Yasuda (2025) followed 39 advanced-level Japanese university students who completed 70 (in-class and at home) 10-minute freewriting activities over an academic semester. Text length significantly increased until the 10th week where it plateaued. Choi and Kim (2025) used a freewriting activity with 149 Korean middle-school students who completed 5-minute writings 8 times over 10 weeks, finding a significant increase in text length, which developed in a non-linear manner. Both Yasuda (2025) and Choi and Kim (2025) found that students made significant gains in other kinds of writing (timed essays in the former study and L2 recount writing in the latter), which suggests that the benefits of freewriting may transfer to other tasks, although it should be noted that both studies lacked control groups. Comparing proficiency levels, Nguyen (2015) found that a group of third-year Vietnamese university EFL students made significant gains in their writing fluency when completing 7-minute timed-writing activities three times a week over a 10-week period, and while no significant changes were observed in a group of first-year students, they outperformed a control group. This study also examined complexity and accuracy, however, very little improvement was found for the former, and none for the latter. Saito (2022) compared fluency development on a 10-minute speed-writing task with Japanese EFL university students of both higher- and lower-intermediate levels. Although the gains were not tested for significance, the results showed that both classes wrote substantially more words over the semester, making average gains of around 18 words for the higher-level group, and 37 words for the lower group. Fluency gains are also reported in a study conducted in a similar context in Herder and Clements (2012).

Several studies have also examined students' responses to completing timed-writing activities, with reactions being largely positive (Choi & Kim, 2025; Hwang, 2010; John, 2019). Hwang (2010) surveyed 8 Thai EFL students, who overwhelmingly rated the activity positively on an 11-item survey, however, the most variation among responses was found in items related to usefulness and confidence. With a class of 40 Indian EFL students John (2019) also found positive open-ended responses, often related to a lack of anxiety due to a reduced focus on accuracy. Choi and Kim (2025) administered a reflection questionnaire in their study of 149 middle-school Korean EFL learners and received 147 responses. They found that the majority of students agreed that freewriting helped them improve their writing skills. In open-ended questions, positive responses also outnumbered negative, with the most common themes being related to confidence, less anxiety about grammar, and engagement during writing. The most common negative points related to limited time, lack of enjoyment, and the difficulty of the task. These results point out that responses to timed writing are generally positive, but they also make clear that this response is not uniform.

While research has shown that timed-writing activities can lead to L2 writing fluency development, there is still very little research examining learners with different proficiency levels, and in particular, the degree to which learners at different levels might prioritize fluency development over other CALF measures. It has been shown that advanced-level learners benefit more from this activity than beginners (Nguyen, 2015), however this finding was based solely on a comparison of initial and final texts, and it is also important to examine the trajectories of fluency development in more detail (Nitta & Baba, 2014). This is because students may be unfamiliar with the task and quickly reach a ceiling effect, and furthermore, research has shown that the development of CALF subcomponents can change in non-linear and interrelated ways (Nitta & Baba, 2014), meaning that growth in fluency might come at the expense of decreases in other measures. While Yasuda (2025) tracked fluency development over 70 writing samples, the study did not examine other CALF measures, making it unclear

how other aspects of writing proficiency are affected when learners are encouraged to focus on text length. Finally, it is important to ascertain student attitudes toward timed writing, as little is known about how university-age learners with different levels of proficiency respond to completing this activity repeatedly over extended periods of time. Therefore, this study investigates fluency development in timed-writing activities over time for learners of two different proficiency levels, focusing on the following research questions:

1. To what extent are lower- and higher-level EFL students able to develop CALF aspects of their writing proficiency in timed-writing activities over the course of an academic semester?
2. To what extent does fluency appear to predominate over other CALF measures when lower- and higher-level EFL students complete timed-writing activities over the course of an academic semester?
3. How do lower- and higher-level EFL students perceive timed-writing activities after weekly practice over an academic semester?

Method

Participants

The study involved two distinct groups from separate private universities with similar populations in Japan, both consisting of non-English majors. One group consisted of 16 first-year students (12 males and 4 females) enrolled in a mandatory reading and writing course, placed in a lower-level class based on scores from English proficiency tests. Additionally, they took compulsory weekly English Discussion and eLearning classes aimed at improving their speaking and listening skills. This class contained 20 students, but as 4 were absent from either the first or final lesson, they were removed from the study. The second group comprised 19 second-year students (11 males and 8 females) from a higher-level academic writing class. Members of this group had applied for entry to this class in their first year of studies which required a score above 600 on a simplified version of the TOEIC test used for university placement, in addition to passing an interview test of communicative English ability. These students took an additional compulsory academic reading class. For both groups, weekly sessions, each lasting 100 minutes, were conducted over a period of 13 weeks. In the first class, a short explanation of the purpose of the study was given, and in the final lesson students filled in a consent form for their data to be used in the analyses. All of the students reported that this was their first experience with timed-writing activities.

Activities

The timed-writing activity focused on in this study aimed to enhance fluency by encouraging learners to write as much as they could in a limited period of time. Topics were jointly decided by the researchers and based on materials that they had used with students of similar levels in previous courses. Familiar topics such as “My favorite season,” “Hobbies,” and “University life” were selected (see Appendix 1 for the list of topics) so that students would be able to continuously write for the full 10 minutes. In the first lesson, the purpose and procedure of the activity was explained, and the teachers showed and explained examples of timed writing to each class. The activity began with a 1-minute brainstorming phase, where students generated ideas without writing or speaking. While the preparation time was quite limited, research has shown that fluency can be positively influenced by even short amounts of planning time (Ellis, 2005). Next, the students wrote for 10 minutes with the instructor recommending that they try to write without stopping, and not to be overly conscious of making errors or editing what they had written. Furthermore, if they encountered difficulties, they were advised to simply start a new sentence, and for lexical issues, use Romanization of Japanese words. While students

wrote, the instructor closely monitored their work and encouraged their efforts. The writing phase concluded with a short peer discussion, where students exchanged ideas about their texts. The remaining 85 minutes of each session featured activities based around individual course objectives. The lower-level students worked on developing both reading and writing skills, integrating reading comprehension tasks, such as identifying main ideas and supporting details, with writing assignments, including two academic essays. The higher-level students' instruction was largely focused on academic writing skills, such as paragraph organization, hedging, and the use of citations, with three academic essays completed over the semester.

Analyses

Each week, the timed writing was entered into text files with the word counts recorded on graphs that were returned to students before beginning the next timed-writing activity. In total, 422 samples were collected, 232 from the higher-level class, and 190 from the lower-level class. Data were not collected from absent students, however, all participants in this study attended at least 9 of the 11 classes between the pre- and post-test, and none were absent for two consecutive weeks. The texts were then analyzed for several CALF measures. Fluency was measured by the total number of words produced during the activity for two reasons. Firstly, due to the large amount of data to be collected during the study, it was determined to be the only practical means of measuring fluency, and secondly to enable comparisons with previously published timed-writing research. Syntactic complexity was measured by the number of clauses per sentence to ascertain the degree of subordination, as well as the average length of sentence. Lexical complexity was measured by the MTLN statistic (obtained using TAALED software (Kyle et al., 2021)), a measure of lexical diversity that is a type-token ratio measure that is less sensitive to text length (Koizumi, 2012). Two accuracy measures were employed in this study: the number of words per verb phrase error and the number of words per syntax error, both having been used in short-term studies of CALF development (Tonkyn, 2012). These measures were selected due to Japanese students' difficulties with both articles and plural forms (associated with errors in noun phrases), as these forms do not have equivalents in their native language and may be less amenable to change over a short-term period. The accuracy and subordination measures were calculated by the researchers. To avoid inconsistencies, five samples of timed writing from a previously taught course were analyzed by each researcher individually and results were compared. Initial results showed that there was 100% agreement with the accuracy measures, and after discussion on how to treat gerund forms, 100% agreement on the subordination measure.

Finally, students were asked to write a comment about the activity in an anonymous end-of-semester survey. The wording of this question was provided in Japanese: *Timed Writing ni tsuite, anata ga kanjiru koto, kanjita koto nado, jiyuu ni kaite kudasai* (please write freely about your thoughts or feelings regarding Timed Writing, including your current or past experiences). The survey was conducted on the last lesson, and students were encouraged to write their answer in Japanese. The results were then translated and sorted into thematic categories. If several themes were mentioned in the same answer, these were categorized into each theme discussed, for example the comment from a lower-level student, "It not only helps with English skills but also helps with daily critical thinking", was classified into both "general English and/or improvement" and "generating content and/or critical thinking skills."

Before answering the research questions, a comparison was made of the first timed-writing samples to determine what, if any, differences could be found in CALF dimensions between the 2 classes. Due to the small sample size, the study lacked sufficient power to calculate reliable effect sizes of between group differences, however, the 95% confidence intervals (CIs) of the means in Week 1 showed that all but one of the CALF measures had non-overlapping

CIs (see Table 1), indicating stable and statistically significant differences (Plonsky, 2015). CIs did overlap for the number of words per verb error (higher-level group 95% CI [40.63, 66.84], lower-level group 95% CI [35.66, 79.11]), however, this overlap was due to several of the lower-level students using only the simple present form (e.g., “I like illumination. It is beautiful”) that reduced the error rate. In contrast, the advanced group used a wider range of verb forms, increasing the possibility for errors (e.g., “We don’t have to turning on such machines”). As will be discussed below, this initial difference in correct verb use was not maintained throughout the course of the study.

Results

To answer research question 1, the degree to which the CALF measures differed between the first and final timed writing for each class, Wilcoxon signed-rank tests were conducted (Table 1). The data were normally distributed, however, because of the small sample sizes, a priori power analyses using G*Power (Faul et al., 2007) were conducted. Power was set at 0.80, a level appropriate in language learning studies (Plonsky, 2015), and the results indicated that the sample size of the higher-level class was sufficient to detect effect sizes of 0.62, and the lower-level class sufficient to detect effect sizes of 0.67. These levels fall just over the benchmark for small effect sizes (0.60) in within-groups L2 research (Plonsky, 2015). Finally, the alpha value was adjusted with a Bonferroni to 0.008 to account for the 6 CALF measures. The results indicated that both groups made significant increases in the number of words written. The higher-level group increased an average of over 40 words, and the standardized effect size index, r , was 0.85, which is between Plonsky’s (2015) L2 benchmarks of 0.60 and 1.00 for small and medium effect sizes in within-groups research. The lower-level group increased an average of slightly below 25 words, with a lower effect size of 0.75. However, as can be seen in Figure 1, considerable differences were observed in both the amount of text produced and the trajectories of fluency development. The average of the higher-level learners trended upwards (with sharp increases between Weeks 1-2 and 4-6) before peaking at time 12, whereas the lower-level group average increased sharply between the first and second week, and thereafter remained fairly stable (the sudden dip in Week 11 appeared to be due to the unexpected difficulty of the social media topic).

Table 1. CALF measure differences: Week 1 and 13

	W1 M	W1 SD	W1 95% CI	W13 M	W13 SD	W13 95% CI	Z	p	r
Fluency									
H-L ⁿ¹	156.05	31.87	[140.69, 171.41]	197.68	47.05	[175.01, 220.36]	-3.72	.000*	.85
L-L ⁿ²	110.37	24.00	[101.70, 119.55]	135.13	29.04	[124.66, 145.59]	-2.98	.003*	.75
Subordination									
H-L ⁿ¹	1.75	.39	[1.56, 1.94]	2.68	.66	[2.36, 2.99]	-3.78	.000*	.87
L-L ⁿ²	1.36	.21	[1.25, 1.47]	1.38	.24	[1.25, 1.51]	-.57	.569	.14
Words per sentence									
H-L ⁿ¹	11.01	2.77	[9.67, 12.35]	14.78	3.06	[13.31, 16.26]	-3.62	.000*	.83
L-L ⁿ²	6.70	1.28	[6.02, 7.38]	7.66	1.10	[7.07, 8.24]	-2.33	.020	.58
Lexical complexity									
H-L ⁿ¹	44.97	11.53	[39.41, 50.53]	50.34	13.20	[43.98, 56.70]	-2.13	.033	.49
L-L ⁿ²	33.21	9.08	[28.37, 38.05]	35.65	8.13	[31.32, 39.98]	-1.45	.148	.36
Words per verb error									
H-L ⁿ¹	53.74	27.19	[40.63, 66.84]	66.51	46.15	[44.27, 88.76]	-.16	.872	.04
L-L ⁿ²	57.39	40.77	[35.66, 79.11]	36.85	24.63	[23.73, 49.98]	-1.34	.189	.34
Words per syntax error									
H-L ⁿ¹	34.13	18.78	[25.08, 43.18]	43.14	16.92	[34.98, 51.30]	-1.24	.215	.28
L-L ⁿ²	18.09	6.64	[14.55, 21.63]	22.00	11.19	[16.04, 27.97]	-.67	.501	.17

Note: H-L = Higher-level class, L-L = lower-level class, n¹ = 19, n² = 16, * = $p < .008$.

Therefore, while a significant difference was found when comparing fluency on the first and final measure, this change appears to be due to the lower-level group's lack of training and familiarity with the activity at the beginning of the semester rather than actual development in their abilities to compose longer texts. Significant change was also found for both of the complexity measures in the higher-level group, with similar effect sizes (0.87 for subordination and 0.83 for the number of words per sentence). No other significant changes were found in any of the other CALF measures, although in the lower-level class, words per sentence increased but fell short of the corrected *p* value. In sum, the higher-level class appeared to increase their fluency and syntactic complexity without any negative effects on lexis or accuracy, but apart from fluency, the measures for the lower-level class did not significantly change over the course of the study. The descriptive statistics for each week of the study can be found in Appendix 2.

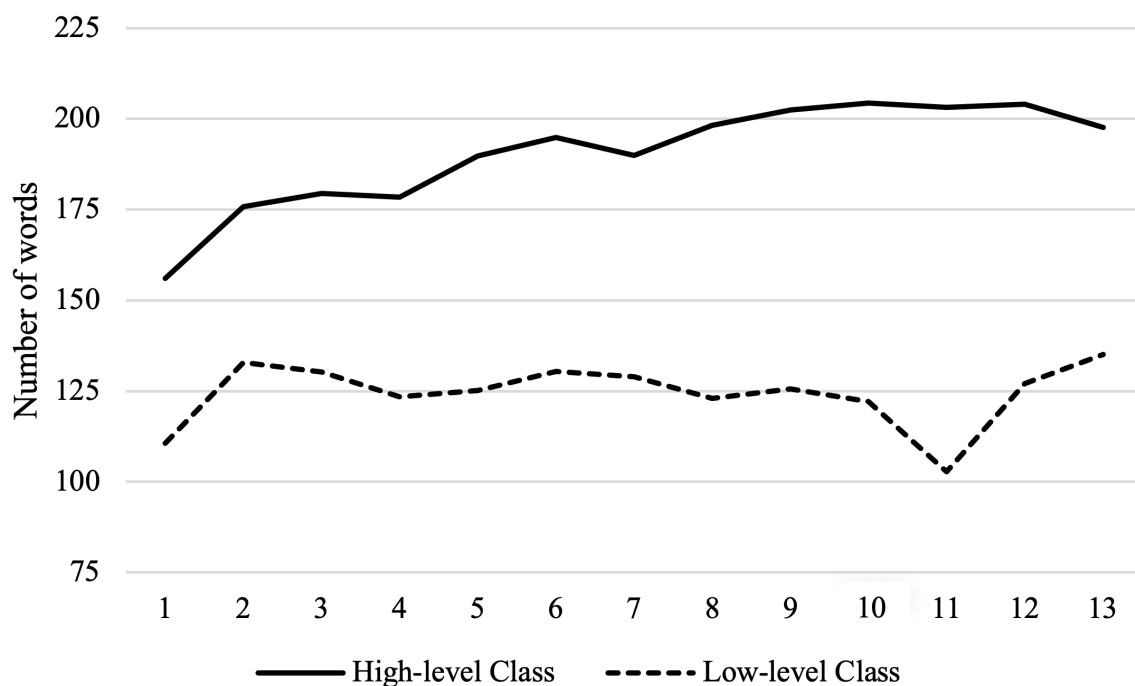


Figure 1. Fluency (as Measured by the Average Number of Words)

Research question 2 examined the degree to which fluency predominated over the other CALF measures to determine if any patterns could be observed that provide insight into whether fluency development might occur at the expense of other CALF dimensions. Following Nitta and Baba (2014), the trajectories of fluency and the other CALF measures were examined by recalculating the data for each variable between 0 and 1 and displaying the results in line graphs (see Verspoor et al., 2011), a procedure that allows for data of differing values to be compared for possible developmental patterns. In these graphs, each value is subtracted from the lowest value and then divided by the highest value, so that 1 represents the highest performance(s) with 0 representing the lowest. While this method of analysis does not include testing for significant interactions, the visualization provides insight into how the CALF measures develop over time in relation to one another. As this study focuses on fluency development, it was of particular interest to see whether higher fluency levels were accompanied by lower values in the other CALF measures because students may have prioritized fluency at the expense of complexity and accuracy. The syntactic complexity measures for the higher-level class showed significant improvement; therefore, they will be discussed in detail. Figures 2a & 2b show the different developmental patterns between fluency and syntactic complexity that

were observed in the two groups. For the higher-level group, fluency and the two complexity measures both initially rose during the first 3 weeks of the study, before following a pattern of gradual increases in fluency, accompanied by fairly stable rates of complexity for the following 8 weeks, with high levels of both complexity and fluency found at the end of the semester. Throughout the study, the levels of both complexity measures were very similar. For the lower-level group, the developmental pattern between fluency and complexity was relatively similar to that of the higher-level group over the first 4 weeks, although fluency slightly decreased in relation to complexity in Weeks 3 and 4. However, there was considerable variation in both complexity measures in the following weeks. In Weeks 4, 8, and 9, longer sentences were accompanied by decreases in fluency, and in Weeks 7, 10, and 13, lower levels of subordination were accompanied by relatively high levels of fluency. Another notable difference in the developmental patterns is that unlike the higher-level group, the two complexity measures occasionally diverged from one another, with longer sentences predominating over subordination. In sum, the higher-level group increased levels of fluency gradually with both measures of complexity remaining relatively stable, while the lower-level group displayed fairly stable levels of fluency with considerably more variation in complexity.

The developmental patterns between fluency and lexical complexity (Figures 3a & 3b) also differed between the groups. The higher-level group occasionally attained high levels of both measures (Weeks 7, 8, and 11) but also had several instances of sudden dips in lexical diversity (Weeks 5, 6, 9, 10, and 12), showing this measure lacked stability over the semester. Similarly, the lower-level class tended toward low levels of lexical diversity (Weeks 4-9) while fluency remained relatively stable, before the two measures converged at high levels in the last two weeks of the semester. Comparing the two groups, it appears that for both, fluency predominated over lexis. However, there were occasions where both groups performed well in both, although the higher-level group did so with more regularity (and at different times) than the lower-level group. Finally, there appeared to be similar developmental patterns between fluency and the accuracy measures for both groups (Figures 4a & 4b). The higher-level group had relatively close levels of accuracy and fluency at the beginning of the semester, but from Week 9 onwards, lower levels of accuracy were fairly consistent as fluency stabilized at a high level. Similarly, the lower-level group had relatively similar levels of fluency and accuracy in the first 2 weeks, but these values began to drop in Week 8 and remained fairly low, even in Week 11 when fluency was at its lowest point in the semester. Therefore, it appeared that the students in both classes gradually prioritized fluency over accuracy (whether consciously or not) as the semester progressed.

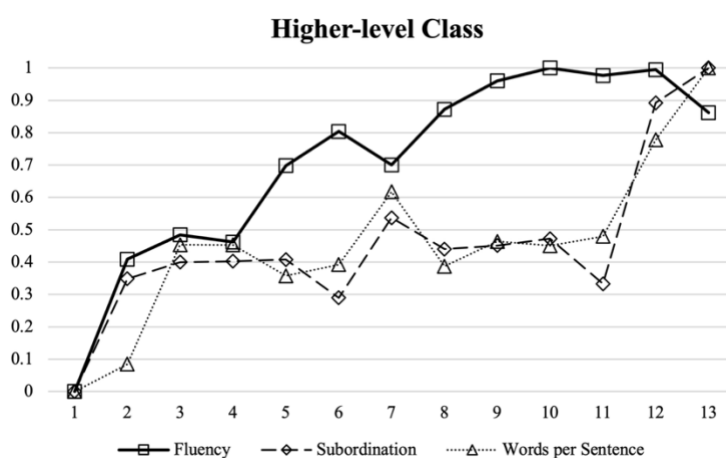


Figure 2a. Normalized Trajectories of Fluency and Complexity (Subordination and Words per Sentence)

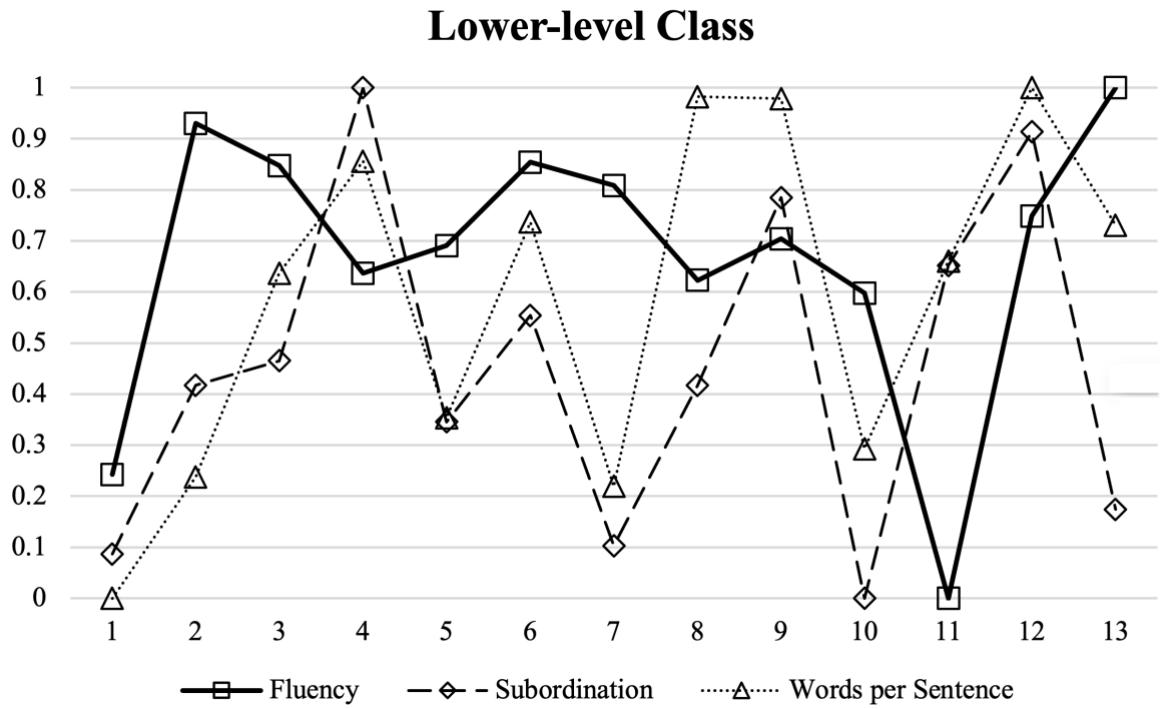


Figure 2b. Normalized Trajectories of Fluency and Complexity (Subordination and Words per Sentence)

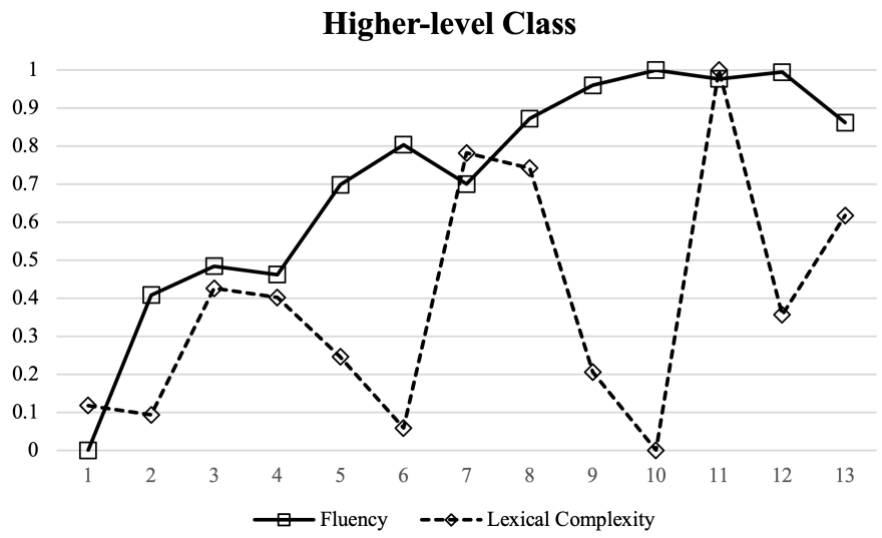


Figure 3a. Normalized Trajectories of Fluency and Lexical Complexity

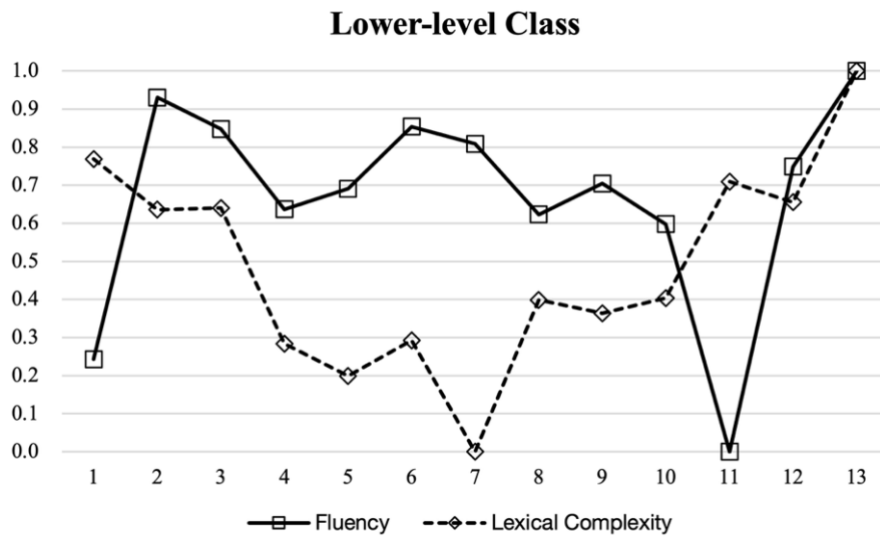


Figure 3b. Normalized Trajectories of Fluency and Lexical Complexity

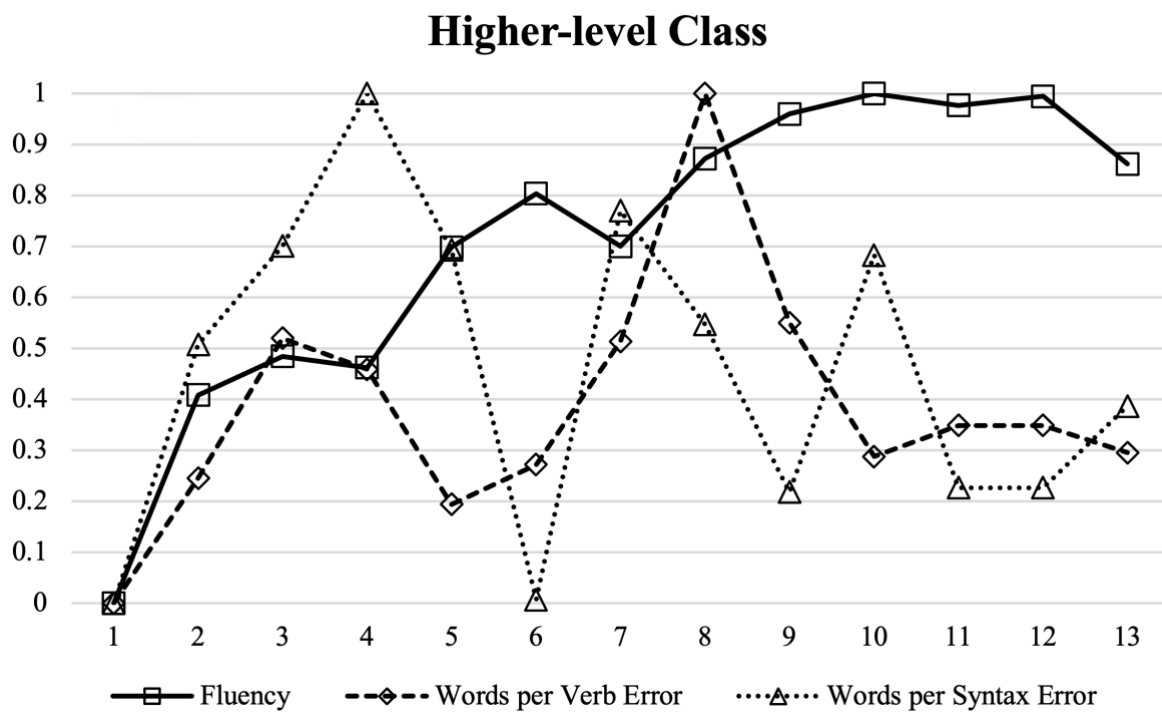


Figure 4a. Normalized Trajectories of Fluency and Accuracy (Number of Words per Verb and Syntax Error)

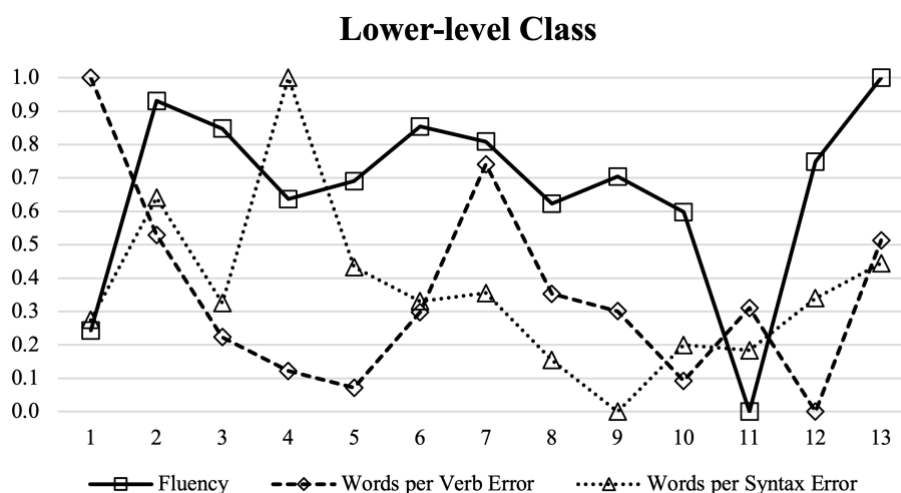


Figure 4b. Normalized Trajectories of Fluency and Accuracy (Number of Words per Verb and Syntax Error)

In terms of perceptions of timed writing, somewhat similar reactions were found in both groups and are summarized in Table 2. 18 students from the higher-level class responded to the open-ended question, whereas 13 students in the lower-level class responded. The most common comments referred to improvement in overall writing ability, writing fluency, and producing content. For example, one student from the higher-level class wrote that the activity “helps me write faster and longer in English”, and a lower-level student wrote “I feel like my ability to write smoothly has gradually improved with practice.” Regarding content and critical thinking, similar opinions were found such as “I’ve started thinking of various ideas related to the topics” from the higher-level group, and “writing about topics I usually don’t think about deeply was fun” from the lower-level group. One major contrast between the two groups was that several students in the higher-level class mentioned opportunities to use grammar and vocabulary skills, with one comment that “timed writing is a good way to practice our grammar and vocabulary that we have mastered or are familiar with.” In contrast, several comments from the lower-level class mentioned difficulties with grammar “I kept getting stuck on grammar, so it was really hard to write a lot” and vocabulary “I sometimes can’t write it in English because I don’t know the words.”

Table 2. Summary of Results from Questionnaire

Aspect of writing	Number of higher-level respondents	Number of lower-level respondents
General English and/or writing improvement	4	5
Writing fluency improvement	7	3
Vocabulary improvement	5	-
Grammar improvement	4	-
Generating content and/or critical thinking	4	4
Motivation/confidence/interest	4	3
Negative points of timed writing	7	3

Another difference found was in how the groups described their positive attitudes toward the activity, with 4 higher-level students mentioning affective issues related to motivation and confidence, whereas 3 students from the lower-level class focused their comments on motivation and their interest in the activity. No students wrote solely negative comments; however, some members of the higher-level class referred to difficulties with content (the need for more specific topics, difficulty writing about their values), while negative comments from the lower-level class mentioned a lack of review, and the effort required to complete the activity. In sum, while the responses from both groups were largely positive, the higher-level class made

more mention of opportunities for improvement in vocabulary and grammar, while the lower-level group focused more on development of their overall English and writing skills.

Discussion

The first research question focused on the degree to which CALF measures increased over the semester and found that both groups were able to make significant gains in fluency, and the higher-level class also made improvements in syntactic complexity. For both groups, lexis and accuracy were unaffected by the activity. Furthermore, the lower-level students' fluency improvement appeared to be due to an initial lack of familiarity with the task because of the sudden increase observed in the second lesson, which was the second highest-value observed in this group over the study. The lower-level group wrote an average of 13.51 words per minute (wpm) in the final activity. Compared with other studies of learners at a similar level, on average these students wrote more than the group in Baba and Nitta (2010), who produced 6.54 wpm, and the lower-level group in Nguyen (2015), who produced 9.42 wpm, but slightly less than those in Choi and Kim (2025), who produced 15.07 wpm. One notable difference between these studies was in how the task was implemented. Baba and Nitta (2010) is unique among these four studies because topics were more sophisticated and students had access to dictionaries. While no gains were found in fluency, significant growth was observed in both syntactic and lexical complexity, a result not seen elsewhere. Comments from student reflections in Baba and Nitta (2014), which also used the same activity, show that learners were concerned with both the content and structure in their writing, which may have influenced the development of other CALF aspects. Comparing these studies suggests that lower-level students may benefit from shorter activities, as the middle-school students in Choi and Kim (2025) were much younger than those in this study but wrote at a similar rate when given a 5-minute freewriting activity.

In contrast, the higher-level group in this study made significant increases in both fluency and complexity. The text length average of 19.77 wpm, however, was considerably lower than the 30.1 wpm seen in Yasuda (2025). Again, the implementation of the activity may have affected this outcome, as students in the latter study were instructed to write “with no regard to grammatical accuracy or logical organization of ideas” (Yasuda, 2025, p. 993), whereas in the present study, the instruction was to *prioritize* content over form. Therefore, the degree to which the quantity of text is emphasized over its quality may affect how much text is produced. Importantly, the students in Yasuda (2025) took 3 more classes of English every week and also completed many more timed-writing activities (70 compared with 13) over an academic semester, which suggests that the number of repetitions might also affect the degree of fluency gains. Regarding complexity, Nitta and Baba (2014) reported significant change in this measure, and similarly, the higher-level group also improved, although the average sentence length was approximately 4 words longer in the present study (while no change in complexity was observed in Nguyen (2015), this may have been due to the complexity measure (the ratio of complex to simple sentences) which was reported by the author as lacking sensitivity).

In sum, the higher-level students in this study were able to write longer texts using more sophisticated grammar, which suggests that learners may need to be above a proficiency threshold in order for timed writing to assist with the development of fluency and complexity. In the context of study abroad, DeKeyser (2007b) has argued that students must have proceduralized some degree of the target language for automatization of language skills to take place, which may account for the larger effect size and differing fluency trajectory of the higher-level group. Reactions to the timed-writing task support this view, as several students from the higher-level class mentioned being able to practice grammar and/or vocabulary that they had previously learned, in contrast to those in the lower-level class who referred to

improvement in their general writing skills, in addition to comments mentioning difficulties with grammar and lexis. Finally, accuracy appears to be an unstable variable in timed writing, although to our knowledge, only one other study (Nguyen, 2015) included an accuracy measurement and also did not observe any significant change. Therefore, while further longitudinal research is needed to learn more about this issue, it appears that accuracy remains unaffected by timed writing, at least over the course of one academic semester.

With regard to research question 2, the lower-level group's level of fluency appeared at times to predominate over complexity (and vice-versa), a pattern also seen in Nitta and Baba (2014). The prioritization of one aspect of language at the expense of another can reflect attentional capacity limitations (Skehan, 2009). For lower-level students, it appears that when they focus on writing longer texts, they often use short and simple sentences; however, when they do use more complex forms, text length is affected. For example, there were unexpected drops in fluency and accuracy on the social media topic (week 11) alongside more complex grammar and lexis, yet this pattern was not observed in the higher-level group. In contrast, when the higher-level group wrote longer texts, their complexity remained stable, and sometimes also increased. With lexical complexity and accuracy, however, both groups struggled to attend to these aspects of their performance while trying to write longer texts, and in terms of accuracy, most of the higher values were observed in the first half of the semester, suggesting that students (following the instructions of the task) gradually prioritized content over form. Whether or not this focus might lead to the automatization of inaccurate language forms as has been cautioned in other studies (Boers, 2014; Thai & Boers, 2016) is unclear, however, these results suggest that classes that use timed writing activities focused on text length also need to provide learners with opportunities to focus on grammatical correctness when writing. In sum, the differences in the developmental patterns of multiple CALF variables suggest that fluency and complexity levels in the higher-level class were consistent across the semester regardless of the topic, whereas in the lower-level class, fluency rose at the expense of other CALF dimensions or was negatively affected when more complex language was used.

The results for research question 3 showed that while both classes reacted positively to the activity, their reasoning was quite different. As mentioned above, several higher-level students reported that they could practice language that they had already learned, one of the main purposes of a fluency activity (Nation, 2008), in addition to comments related to increases in writing confidence. These kinds of comments were not seen in responses from the lower-level class, supporting the interpretation that students may need to pass a proficiency threshold in order to improve writing fluency in timed-writing activities. However, responses from both classes regarding content and critical thinking described the activity as not only an opportunity to practice language, but also to develop their thinking skills and ideas, showing that for some, the activity was valued for reasons beyond foreign language learning. While these positive responses are similar to those found in prior studies (Choi & Kim, 2025; Hwang, 2010; John 2019), there are also several notable differences. The largest similarities were that the majority of students responded positively to the activity and felt that their writing skills improved. However, in contrast to previous studies, no mention was found of reduced anxiety due to less focus on grammar, which was reported in previous studies (Choi & Kim, 2025; John, 2019). This may be related to how the activity was presented to the classes, as instructions were to prioritize meaning over form, whereas other studies have more strongly instructed students to continue writing without regard for accuracy (Choi & Kim, 2025; Yasuda, 2025). Finally, in contrast to the reflections reported in Baba and Nitta (2014), no mention was made of text organization, which is likely due to the easier topics and focus on text length in the current study. These comparisons with prior studies suggest that while students react positively to

timed-writing activities, the manner of presentation and instruction affects perceptions of the task.

Pedagogical Implications

These findings suggest that, in terms of language proficiency, the activity was at a reasonably good level for the higher-level students, as significant development was made in their writing fluency. While the lower-level learners' gains were less impressive, it should be emphasized that these data reflect short-term development only, and in the long term more balanced growth might be seen for this group. To date, very little research has examined the texts produced by higher- and lower-level learners completing timed-writing activities over time, and several pedagogical implications for fluency development or writing skills can be inferred from the findings. First, it may be helpful for teachers to measure levels of fluency and other CALF measures for several weeks to determine how well the class is able to attend to multiple dimensions of their performance. If students are not able to perform relatively consistently in fluency and complexity, the activity may need to be adjusted to encourage use of familiar language which facilitates the development of automatization processes (DeKeyser, 2007b). One option for lower-level students would be to keep the time allotted to writing relatively short, as gains have been seen with lower-level students writing for 5 minutes (Choi & Kim, 2025). Second, lower-level learners may benefit from additional support built into the task. While previous studies have shown that repetition of the same task does not always result in more fluent written texts (Hosoda, 2018; Nitta & Baba, 2014), lower-level learners might benefit from topics based on class materials that they have previously discussed or written about in order to facilitate attention to more aspects of their language while writing. Another possibility is to simplify the task by having students write shorter answers to more specific questions. Finally, higher-level students might benefit from post-writing activities focused on accuracy, due to the lack of improvement in this area over the course of the study. While there is the possibility that focus-on-form activities might negatively affect fluency, one possibility would be for students to check their work together with a partner and identify specific aspects that could be improved. Alternatively, learners could be given the opportunity to rewrite their text as a homework activity, keeping a record of the changes made to content, structure, and lexis, with the aim of building awareness of typical errors. It is hoped that future studies exploring how timed-writing tasks might be adapted to suit learners of different proficiency levels and backgrounds will be conducted.

Conclusion

This study compared the development of CALF measures on a timed-writing task between higher- and lower-level EFL learners over the course of an academic semester, finding that while both groups made improvements in fluency, the higher-level class made significant improvements that developed across the semester, in contrast to the lower-level class whose gains appeared to be due to an initial lack of familiarity with the task. Furthermore, analyses of the normalized trajectories of fluency and the other CALF measures showed that the higher-level students were more able to attend to syntactic complexity and, occasionally, lexis while writing longer texts, whereas the lower-level group's performances more often prioritized fluency over complexity, or vice versa. Finally, accuracy levels fluctuated for both groups over the course of the study, suggesting that even relatively higher-level learners are not able to attend to both fluency and accuracy in their writing when under time-pressured conditions to produce text. Therefore, while timed-writing activities do not appear to be sufficient for developing all aspects of writing proficiency, it seems that they can play an important role in enabling students to produce longer texts in shorter periods of time, provided that learners have attained a suitable level of language proficiency.

It should be emphasized that there are several limitations of this study. First, the lack of control groups means the changes in writing fluency may have resulted from activities outside of the timed writing, although students from both classes reported that they did little writing in their other English classes. Second, because the timed writing was an unassessed weekly activity, it remains unclear whether the language that they produced is indicative of their actual writing proficiency. Furthermore, the developmental trajectory analyses looked at overall group trends in order to determine how performance on the timed-writing activity developed at the class level. With regard to L2 writing development, group patterns of change have been shown to differ from individual patterns (Larsen-Freeman, 2016). Finally, the study did not include a measure of functional adequacy (Pallotti, 2009) or other qualitative analyses that may have provided insight into the content and organization of student texts. Issues related to individual development and text quality will be addressed in a follow-up study comparing individual writers who made gains in their fluency with those who did not. Nevertheless, it is hoped that the examination of proficiency level and fluency development presented here will contribute to the small but growing number of studies examining timed writing in instructed contexts.

About the Authors

Timothy Doe is a Senior Assistant Professor at the School of Information and Communication at Meiji University in Tokyo. His research interests include the teaching and assessment of speaking skills, curriculum and materials development, and fluency development in speaking and writing activities. ORCID ID: 0000-0002-6675-1961

Teruyo Nakao is an assistant professor at Toyo Gakuen University in Tokyo, Japan. Her research interests include teacher development, collaborative learning, task-based learning, and learner autonomy. She is currently conducting research on the relationship between writing fluency activities and vocabulary development. ORCID ID: 0009-0008-1557-9939

To Cite this Article

Doe, T. & Nakao, T. (2026). Fluency development in timed writing: The role of proficiency. *Teaching English as a Second Language Electronic Journal (TESL-EJ)*, 30(1). <https://doi.org/10.55593/ej.30117a7>

References

- Abdel Latif, M. M. (2013). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics*, 34(1), 99–105. <https://doi.org/10.1093/applin/ams073>
- Abdi Tabari, M., & Golparvar, S. E. (2024). The interplay of task repetition and task rehearsal in L2 written production across varied proficiency levels. *Language Teaching Research*. Epub ahead of print 1 August 2024. <https://doi.org/10.1177/13621688241266940>
- Baba, K., & Nitta, R. (2010). Dynamic effects of task type practice on the Japanese EFL university student's writing: Text Analysis with Coh-Metrix. *The Florida AI Research Society*.
- Baba, K., & Nitta, R. (2014). Phase transitions in development of writing fluency: From a complex dynamic systems perspective. *Language Learning*, 71(1), 31-69. <https://doi.org/10.1111/lang.12033>
- Baba, K., & Nitta, R. (2021). Emergence of multiple groups of learners with different writing-development trajectories in classroom: Growth mixture modeling. *Journal of Second Language Writing*, 54, 100856. <https://doi.org/10.1016/j.jslw.2021.100856>

- Boers, F. (2014). A reappraisal of the 4/3/2 activity. *RELC Journal*, 45, 221-235. <https://doi.org/10.1177/0033688214546964>
- Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks* (pp. 23-48). Longman.
- Bygate, M. (2018). Introduction. In M. Bygate (Ed.), *Language learning through task repetition* (pp. 1-25). Benjamins.
- Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written communication*, 18(1), 80-98. <https://doi.org/10.1177/0741088301018001>
- Choi, Y. S., & Kim, M. (2025). The effects of freewriting on L2 writing fluency, emotions, and perceptions among secondary EFL learners. *International Journal of Applied Linguistics*. Epub ahead of print 23 October 2025. <https://doi.org/10.1111/ijal.70026>
- De Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning*, 46(3), 529-555. <https://doi.org/10.1111/j.1467-1770.1996.tb01246.x>
- DeKeyser, D. (2007a). Introduction: Situating the concept of practice. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 1-18). Cambridge University Press.
- DeKeyser, D. (2007b). Study abroad as foreign language practice. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 208-226). Cambridge University Press.
- Ellis, R. (2005). Planning and task-based performance: Theory and research. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 3-34). Benjamins.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191. <https://doi.org/10.3758/BF03193146>
- Fukunaga, T. (2023). L2 writing development through two types of writing task repetition. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1109-1138. <https://doi.org/10.1515/iral-2021-0144>
- Herder, S., & Clements, P. (2012). Extensive writing: A fluency-first approach to EFL writing. In T. Muller, S. Herder, J. Adamson, & P. Shigeo Brown (Eds.), *Innovating EFL teaching in Asia* (pp. 232-244). Palgrave.
- Hinkel, E. (2004). *Teaching academic ESL writing*. Routledge.
- Hosoda, N. (2018). Effects of speedwriting and task repetition on the development of writing fluency. *Language Media and Learning Research*, 27, 47-65. <https://www.kandagaigo.ac.jp/kuis/cms/wp-content/uploads/2018/04/02.pdf>
- Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy, and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy, and fluency in SLA* (pp. 1-20). Benjamins.
- Hwang, J. A. (2010). A case study of the influence of freewriting on writing fluency and confidence of EFL college-level students. *Second Language Studies*, 28(2), pp. 97-134. <http://hdl.handle.net/10125/40706>

- John, D. (2019). 'Free Writing' versus 'Writing Fluency'. *Journal of Asia TEFL*, 16(1), 369. <https://doi.org/10.18823/asiatefl.2019.16.1.26.369>
- Kim, Y., Kang, S. Hyunae, Y., Kim, B., & Choi, B. (2020). The role of task repetition in a Korean as a foreign language classroom: writing quality, attention to form, and learning of Korean grammar. *Foreign Language Annals*, 53(4), 827-849. <https://doi.org/10.1111/flan.12501>
- Koizumi, R. (2012). Relationships between text length and lexical diversity measures: Can we use short texts of less than 100 tokens? *Vocabulary Learning and Instruction*, 1(1), 60-69. <https://doi.org/10.7820/vli.v01.1.koizumi>
- Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity using direct judgements. *Language Assessment Quarterly* 18(2), pp. 154-170. <https://doi.org/10.1080/15434303.2020.1844205>
- Larsen-Freeman, D. (2016). Classroom-oriented research from a complex systems perspective. *Studies in Second Language Learning and Teaching* 6(3), pp. 377-393. <https://doi.org/10.14746/ssl.t.2016.6.3.2>
- Macalister, J., & Nation I. S. P. (2020). *Language curriculum design* (2nd ed.). Routledge.
- Min, J. (2023). Writing fluency instructional practices for college-level multilinguals in a US intensive writing course: From an activity theory perspective. *Language Teaching Research*. Epub ahead of print 21 January 2023. <https://doi.org/10.1177/13621688221146089>
- Nation, I. S. P. (2008). *Teaching ESL/EFL reading and writing*. Routledge.
- Nguyen, L. T. C. (2015). Written fluency improvement in a foreign language: The effects of extensive writing and task repetition. *TESOL Journal*, 6(3), 447-472. <https://doi.org/10.1002/tesj.186>
- Nitta, R., & Baba, K. (2014). Task repetition and L2 writing development. In H. Byrnes and R. M. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 107-136). Benjamins.
- Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30(4), 590-601. <https://doi.org/10.1093/applin/amp045>
- Penn, S., & Lim, H. (2016). The effects of freewriting exercises on adult Korean students' English learning. *Journal of Asia TEFL*, 13(4), 313. <https://doi.org/10.18823/asiatefl.2016.13.4.5.313>
- Plonsky, L. (2015). Statistical power, *p* values, descriptive statistics, and effect sizes: A back to basics approach to advancing quantitative methods in L2 research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 23-45). Routledge.
- Saito, Y. (2022). Exploring the potential role of speed writing activities in academic writing courses. *Dokkyo Journal of Language Teaching and Learning* 11, 59-73. <https://dokkyo.repo.nii.ac.jp/record/3220/files/P-095-D83gc-11-3.pdf>
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510-532. <https://doi.org/10.1093/applin/amp047>
- Skehan, P., & Foster, P. (2012). Complexity, accuracy, fluency and lexis in task-based performance: A synthesis of the Ealing research. In A. Housen, F. Kuiken, & I. Vedder

(Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 199-220). Benjamins.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16(3), 371-391. <https://doi.org/10.1093/applin/16.3.371>

Tavakoli, P. (2025). Assessment of second language fluency. *Language Teaching*, 58(3), 312–328. <https://doi.org/10.1017/S0261444824000417>

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–277). Benjamins.

Thai, C. & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*, 50, 369-393. <https://doi.org/10.1002/tesq.232>

Tonkyn, A. (2012). Measuring and perceiving changes in oral complexity, accuracy and fluency: examining instructed learners' short-term gains. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: complexity, accuracy, and fluency in SLA*. Benjamins.

Verspoor, M., Lowie, van Geert, P., van Dijk, & Schmid, M. S. (2011). How to sections. In M. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development*. Benjamins.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. University of Hawaii Press.

Yasuda, R. (2025). Fluency development through freewriting and transfer to other more structured tasks. *Language Teaching Research*, 29(3), 986-1006. <https://doi.org/10.1177/13621688221084899>

Appendices

Appendix 1. Timed-writing Topics

Week	Topic
1	My favorite season
2	Hobbies
3	Sunday
4	Sports
5	Traveling
6	Games
7	Food
8	Friends
9	Money
10	Family
11	Social media
12	Study
13	University life

Appendix 2. Means and Standard Deviations for the CALF Measures

<i>Higher-level Class</i>						
	Fluency	Subordination	Words per sentence	Lexical Complexity	Words per verb error	Words per syntax error
Week 1	156.05 (31.87)	1.75 (.39)	11.01 (2.77)	44.97 (11.53)	53.74 (24.77)	34.13 (18.78)
Week 2	175.79 (38.00)	2.07 (.49)	11.33 (2.07)	44.70 (11.07)	62.4 (47.98)	52.35 (25.81)
Week 3	179.44 (33.01)	2.12 (.51)	12.72 (3.14)	48.28 (14.15)	72.08 (49.09)	59.26 (34.78)
Week 4	178.38 (37.48)	2.13 (.65)	12.36 (3.60)	48.02 (14.06)	69.93 (44.62)	70.0 (48.92)
Week 5	189.78 (40.10)	2.02 (.46)	12.49 (2.42)	46.34 (13.04)	60.58 (35.51)	59.01 (27.90)
Week 6	194.88 (45.10)	2.25 (.53)	13.34 (2.89)	44.34 (13.70)	63.33 (54.70)	34.37 (11.26)
Week 7	189.9 (40.93)	2.16 (.48)	12.47 (2.82)	52.11 (18.57)	71.87 (55.10)	61.76 (55.56)
Week 8	198.19 (47.21)	2.17 (.41)	12.76 (1.95)	51.68 (14.24)	89.02 (56.63)	53.75 (38.10)
Week 9	202.42 (36.79)	2.28 (.52)	13.18 (2.61)	45.91 (10.80)	73.14 (40.43)	41.94 (16.72)
Week 10	204.35 (45.12)	2.19 (.96)	12.71 (4.94)	43.7 (12.78)	63.9 (55.25)	58.64 (52.12)
Week 11	203.24 (38.80)	2.06 (.40)	12.82 (2.82)	54.45 (15.39)	66.04 (50.58)	42.26 (26.64)
Week 12	204.12 (45.24)	2.58 (.55)	13.95 (3.03)	47.53 (12.99)	64.15 (46.22)	48.01 (22.01)
Week 13	197.68 (47.05)	2.68 (.66)	14.78 (3.06)	50.34 (13.20)	66.51 (46.15)	43.14 (16.93)
<i>Lower-level Class</i>						
	Fluency	Subordination	Words per sentence	Lexical Complexity	Words per verb error	Words per syntax error
Week 1	110.63 (24.77)	1.36 (.21)	6.70 (1.28)	33.21 (9.01)	57.39 (40.77)	18.09 (6.64)
Week 2	132.88 (34.56)	1.45 (.25)	7.01 (1.04)	31.81 (7.69)	37.41 (35.51)	26.53 (17.63)
Week 3	130.20 (32.43)	1.46 (.22)	7.53 (1.37)	31.85 (9.27)	26.78 (13.93)	19.26 (6.61)
Week 4	123.38 (28.00)	1.60 (.29)	7.82 (1.27)	28.09 (8.17)	23.26 (17.26)	34.81 (18.34)
Week 5	125.13 (27.71)	1.43 (.37)	7.16 (1.46)	27.20 (8.37)	21.52 (11.01)	21.74 (8.71)
Week 6	130.40 (33.23)	1.48 (.37)	7.66 (1.53)	28.18 (6.81)	29.38 (12.85)	19.38 (9.73)
Week 7	128.94 (32.49)	1.36 (.21)	6.99 (1.15)	25.10 (4.76)	44.74 (45.12)	19.96 (9.42)
Week 8	122.93 (35.04)	1.45 (.37)	7.99 (2.53)	29.31 (6.89)	31.30 (16.22)	15.32 (6.46)
Week 9	125.56 (34.69)	1.54 (.33)	7.98 (1.69)	28.94 (5.27)	29.49 (14.41)	11.76 (3.82)
Week 10	122.11 (28.34)	1.33 (.17)	7.08 (0.65)	29.36 (5.62)	22.24 (13.42)	16.37 (6.65)
Week 11	102.77 (27.32)	1.51 (.27)	7.56 (1.41)	32.59 (9.05)	29.81 (16.07)	15.99 (6.43)
Week 12	127.00 (35.66)	1.58 (.35)	8.01 (1.45)	32.01 (11.58)	19.04 (7.51)	19.60 (13.46)
Week 13	135.13 (29.04)	1.38 (.24)	7.66 (1.10)	35.65 (8.13)	36.85 (24.63)	22.00 (11.19)

Copyright of articles rests with the authors. Please cite TESL-EJ appropriately.