A CAF Perspective on L2 Development Through Writing Practice: An Exploratory Study

J. Paul Marlowe
Kwansei Gakuin University, School of Policy Studies
Mayumi Asaba
Kyoto Sangyo University, Faculty of Foreign Studies

Reference Data

Marlowe, J. P., & Asaba, M. (2022). A CAF perspective on L2 development through writing practice: An exploratory study. In P. Ferguson, & R. Derrah (Eds.), *Reflections and new perspectives*. JALT. https://doi.org/10.37546/JALTPCP2021-15

Although complexity, accuracy, and fluency (CAF) are considered important aspects of language proficiency, little is known about the role of different writing activities in developing components of L2 writing proficiency. Therefore, three groups who received regular targeted writing practice in either sentence-combining, translation, or timed-writing were compared over the course of one academic semester on measures of CAF. Three writing samples from each group were collected and statistically analyzed for quantitative measures of CAF. Furthermore, interviews were conducted with participants from each group to explore their perceptions of the tasks and the effect that tasks have on participants' L2 writing effect on their L2 writing. The results of this exploratory study indicate that sentence-combining aids growth in complexity and that the timed-writing group significantly outperformed other groups in fluency, while growth in these areas seemed to diminish accuracy for all groups. These findings were confirmed by the qualitative data with implications for the use of these tasks for developing writing proficiency.

複雑さ、正確さ、流暢さ（CAF）は言語能力を構成する重要な要素である。しかし、この3要素を習得させるライティングの練習方法についてはほとんど知られていない。本研究では、１学期間に渡り３種類の練習（センテンス・コンバイニング、翻訳、時間制限を設けたライティング）を行ったグループのライティングサンプルを、CAFを使って統計学的に分析・比較した。さらに、インタビューを通じて、練習に関する学生の意見と、練習がどのような影響をライティングに与えると学生自身が認識しているのか、という２点について調査した。結果として、全てのグループにおいて正確さは減少し、センテンス・コンバイニング練習は複雑さを、そして時間制限を設けたライティング練習は流暢さを改善することが判明した。インタビュー・データの結果は統計分析結果を裏付け、さらに３種類の練習方法をどのように使用すべきかという点も示唆した。

The development of L2 proficiency is considered multicomponential and the components are often captured using indices of complexity, accuracy, and fluency (CAF), with lexis being included more recently as an independent measure of complexity and collectively referred to as CALF (Housen et al., 2012; Skehan, 2018). Ortega (2003) defined complexity as "the range of forms that surface in language production and the degree of sophistication of such forms" (p. 492). According to Wolfe-Quintero, et al. (1998), accuracy is defined as "the ability to be free from errors while using language to communicate" (p. 33), and fluency is defined as "the number of words or structural units a writer is able to include in their writing within a particular period of time" (p. 14). These three constructs are often used as variables to measure learner performance and gauge L2 development in the writing classroom (Kuiken et al., 2010; Wigglesworth & Storch, 2009).

From the perspective of CAF, the goal of classroom writing pedagogy is to help learners make more complex, accurate, and fluent writing. Therefore, it is essential to understand what types of pedagogical interventions can help learners achieve proficiency through the development of these components. In a dissertation study, Marlowe (2019) investigated the effect of different writing practice tasks on the development of complexity in Japanese university student writing. In the results, Marlowe reported that learners practicing sentence-combining showed significant gains over two other groups that practiced translation and timed-writing on two of the five complexity measures (mean length of sentence and T-units per sentence) which indicated that the writers in the sentence-combining group used more coordination in their sentence structures. Furthermore, because all groups made gains in complexity across time but scored lowest on accuracy by raters, Marlowe also suggested that there might have been a tradeoff between the development of complexity at the cost of accuracy. This finding is in line

with tradeoff theories reported by Ellis (1994) and developmental sequence patterns proposed by Housen et al. (2012), in which complexity development precedes accuracy.

To this day, it is still unclear how CAF components are interconnected and how they develop over time (see Skehan, 2018, for a recent review). CAF is regarded as part of a system (Housen & Kuiken, 2009; Larsen-Freeman, 2009) that captures developmental trends, often revealing a dynamic relationship between change and interaction (Norris & Ortega, 2009). Another reason the relationship between the components remains unclear is because there are many factors that affect the development of CAF, including learner variables, the type of input, and the complexity of tasks (Housen & Kuiken, 2009; Robinson, 2005). Finally, CAF related research rarely includes qualitative data in the form of students' perspectives on their own development.

Therefore, given the lack of research on the effect of practice tasks on L2 writing development, the still unclear relationship between CAF variables in L2 writing development, and the dearth of qualitative data in CAF studies, the researchers in this study seek to address the following questions:

RQ1. What is the differential impact of translation, sentence-combining, and timed-writing on learners' complexity, accuracy, and fluency in writing over time?

RQ2. What is the relationship between CAF variables as learners' writing develops over time?

RQ3. How do learners perceive the effect of writing practice on their own development?

Study Design

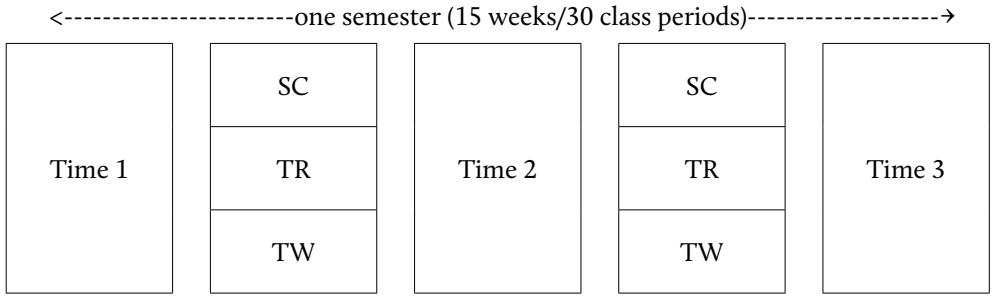
The study was exploratory in nature and was conducted using a mixed-methods explanatory and concurrent design, meaning that quantitative and qualitative data were gathered simultaneously, and the qualitative data was used to help explain the quantitative results (Creswell, 2009). Participants were 43 first-year students at a private university in Western Japan. The students consisted of 18 females and 25 males with an average TOEIC (L&R) score of 403 (*SD* = 96.19). The students were enrolled in three classes of a first-year writing course taught by three different instructors, one of which was the first author. The three classes were streamed into three levels based on an initial TOEIC test at the beginning of the spring semester and then adjusted for the fall semester using English grades. The timed-writing group was the highest level, followed by the translation group, and then the sentence-combining group. All groups followed the same first-year writing curriculum, which included basic academic writing skills with a focus on paragraphs and process writing. At the outset of the study, all participants who agreed to participate signed an informed consent form which included a detailed explanation of the study in Japanese.

Each of the three intact classes were randomly assigned either practice in sentence-combining (n = 18), translation (n = 16), or timed-writing (n = 9). Sentence-combining was chosen because there is evidence to indicate that practicing sentence-combining can hasten the development of L2 writing complexity (Marlowe, 2016; 2019). Translation, a common practice in East Asian L2 education, can help learners focus more on lexical aspects of sentence writing (Marlowe & Asaba, 2017) and can be used as an effective means of teaching grammatical accuracy (Källkvist, 2008). Finally, studies on fluency writing (Baba, 2020; Baba & Nitta, 2014) provide evidence that learners practicing timed writing improve their writing fluency over time.

Data collection took place during the fall semester, which consisted of two class meetings a week for 15 weeks for a total of 30 classes. Writing samples were collected three times during the semester: during the first week (Time 1), the eighth week (Time 2), and the last week (Time 3). In between, the participants completed their respective writing practice at the beginning of each class meeting (Figure 1).

Figure 1

Research Design Overview



Note: SC = sentence–combining group; TR = translation group; TW = timed-writing group.

Procedures

The participants in this study were assigned one of three different types of writing practice tasks. Here, we refer to writing practice tasks as exercises that are designed to improve specific writing skills that can be applied to a variety of writing composition texts. The tasks were allotted the first 20 minutes of each class period. The sentence-combining and translation groups received worksheets at the start of each class with the same target structure and the same target model answers. The worksheets were all developed by the first author following a sequence of target structures recommended by Cooper (1973) and Lawlor (1983) that included coordinates, adverbials, and noun modifiers. One target structure was presented each week with the first class focused on sentence level exercises and the second class on paragraphs. The third intact group practiced timed writing by selecting topics from a list and writing as much as they could on the topic without stopping or using other resources, such as a dictionary, for 10 minutes. Then they counted the words and recorded it on a graph which was collected by the teacher each week. The following are examples of each task:

1. **Translation task**: Translate the following sentence from Japanese to English.
   クリスは旅行することと新しい人々に会うことが好きだが、飛行機に乗るのが嫌いだ。
2. **Sentence-combining task**: Combine the following short sentences into one longer sentence.
   Chris likes traveling. Chris likes meeting new people. Chris hates flying.
3. **Timed-writing task**: Choose a topic and write non-stop for ten minutes.

Two writing samples were collected via Google Forms during each test time. The writing prompts were narrative topics on past experiences such as birthdays, vacations, first day of school, and best trip. Narrative topics were chosen because they are personalized, familiar to all participants, and appropriate for first-year college level L2 learners (Byrnes et al., 2010). Participants were given 15 minutes to write freely on two of the topics at each test time for a total time of 30 minutes. Two topics were given in order to account for the effect of topic on the CAF variables (Lu, 2011; Yoon, 2017).

Measures and Analyses

Repeated measures analysis of variance (ANOVA) was used to analyze the variables across time and one-way ANOVA at each test time was used to examine between group differences at each test time and on each of the measures (see Table 1). Multiple ANOVAs were used instead of a Multiple Analysis of Variance (MANOVA) because the small and unequal sample sizes can diminish the power of a MANOVA (Tabachnick & Fidell, 2014). Furthermore, to understand the impact of the treatments, effect sizes (*d*) were calculated using Plonsky and Oswald's (2014) guidelines for small (0.4), medium (0.7), and large (1.0) between group contrasts.

Complexity was measured with three indices, in line with the recommendations of Norris and Ortega (2009), who argued that the syntactic complexity should include measures that target at least three constructs. Therefore, mean length of sentence (MLS) was chosen as a measure of general complexity, clauses per T-unit (C/T) was selected to target subordination, and T-units per sentence (T/S) were used to target low proficiency learners' use of coordination. T-units are "terminal units" and consist of an independent clause and any clauses attached to it (Hunt, 1965). Complexity measures of the writing samples were analyzed using the free online L2 Syntactical Complexity Analyzer (L2SCA) developed by Lu (2010).

Accuracy was measured using error-free clause ratio (EFCR) following guidelines established by Polio (1997). The clause was used as the unit of analysis because it is considered a more precise and appropriate unit for beginner and low-intermediate learners (Ishikawa, 1995). Average rating scores were calculated from evaluations by both researchers in order to improve the accuracy and to provide more precise judgments.

Finally, fluency was measured based on the total number of words written in the allotted 15-minute time period (W/15M).

Table 1

Measures of CAF

Category	Measures	Calculation
Complexity	Mean Length of Sentence (MLS)	Total number of words divided by number of sentences
	Clauses per T-units (C/T)	Total number of clauses divided by number of T-units
	T-units per sentence (T/S)	Total number of T-units divided by number of sentences
Accuracy	Error-free clause ratio (EFCR)	Total number of error free clauses divided by number of clauses
Fluency	Words per 15 minutes (W/15M)	Average number of words written in 15 minutes

Results

The results of the analyses were analyzed using the Statistical Package for Social Science (SPSS) version 28.01 software program. Comparisons were made across times and between groups. Complete descriptive statistics for all measures, groups, and test times are available in the Appendix. It was hypothesized that practicing sentence-combining would have the greatest effect on measures of complexity, practicing translation would have the greatest effect on accuracy, and practicing timed-writing would have the greatest effect on fluency.

The descriptive statistics for the translation group are reported in Table 2. Although the means for the complexity measures rose slightly from Time 1 to Time 2, and then decreased at Time 3, there were no significant differences between time on any of the measures for the translation group. On accuracy, the means decreased between all test times and with a statistically significant effect of time on EFCR, $F(1, 15) = 8.025$, $p < .01$. Post hoc tests revealed that there was a significant difference between Time 1 and Time 3 with a medium to large effect size ($d = .83$). The means for the fluency measure increased across each test time with a statistically significant effect of time on W/15M, $F(1, 15) = 23.573$, $p < .001$. Post hoc tests revealed significant differences and large effect sizes between Time 1 and 3 ($d = 1.43$) and Time 2 and 3 ($d = .95$).

Table 2

Descriptive Statistics for the Translation Group

Measure		Time 1		Time 2		Time 3	
		SD	*M*	*SD*	*M*	*SD*	*M*
Complexity	MLS	9.06	2.63	10.34	2.23	10.00	2.88
	C/T	1.22	.18	1.36	.18	1.32	.15
	T/S	1.05	.13	1.11	.11	1.10	.07
Accuracy	EFCR	.69	.19	.61	.18	.54	.17
Fluency	W/15M	71.47	29.28	85.00	32.06	115.97	32.97

The results of repeated measures ANOVA for the sentence-combining group revealed significant effects for time. The descriptive statistics are reported in Table 3. The means for the complexity measures rose across time, with statistically significant differences for MLS, $F(2, 20) = 20.498$, $p < .001$. Post hoc tests revealed significant differences and medium to large effect sizes between Times 1 and 2 ($d = .95$) and Times 1 and 3 ($d = 1.04$). There was statistically significant effect of time for C/T, $F(1, 17) = 10.213$, $p < .01$, with post hoc results revealing significant differences and a medium to large effect size between Time 1 and Time 2 ($d = .81$), and Time 1 and 3 ($d = .93$). There was also a statistically significant effect of time on T/S, $F(1, 17) = 9.295$, $p < .001$, with post hoc results revealing significant differences and a large effect sizes between Time 1 and Time 2 ($d = 1.12$), and Time 1 and 3 ($d = 1.17$). There were no statistically significant effects for time on the measures for accuracy and fluency.

Table 3

Descriptive Statistics for the Sentence-combining Group

Measure		Time 1		Time 2		Time 3	
		SD	*M*	*SD*	*M*	*SD*	*M*
Complexity	MLS	8.12	2.22	10.86	3.42	11.02	3.25
	C/T	1.19	.18	1.38	.28	1.34	.14
	T/S	1.03	.07	1.14	.12	1.16	.14
Accuracy	EFCR	.66	.14	.61	.17	.59	.17
Fluency	W/15M	72.92	27.61	69.75	22.34	81.50	32.03

The descriptive statistics for the timed-writing group are reported in Table 4. Whereas the means for both T/S slightly increased across the three times, the means for MLS and C/T rose from Time 1 to Time 2, and then decreased at Time 3. There were no statistically significant differences between any of the times on any of the complexity measures. The means for EFCR decreased slightly across time, but there was no statistically significant differences. For fluency, the means increased sharply across all three times with a statistically significant effect of time on W/15M, $F(1, 8) = 15.105$, $p < .001$. Post hoc tests revealed significant differences and large effect sizes between Times 1 and 3 ($d = 1.66$) and Times 2 and 3 ($d = 1.00$).

Table 4

Descriptive Statistics for the Timed-writing Group

Measure		Time 1		Time 2		Time 3	
		SD	*M*	*SD*	*M*	*SD*	*M*
Complexity	MLS	10.21	2.35	11.94	2.60	11.75	2.64
	C/T	1.40	.25	1.51	.20	1.44	.20
	T/S	1.10	.05	1.12	.08	1.14	.10
Accuracy	EFCR	.65	.12	.64	.14	.58	.11
Fluency	W/15M	116.44	35.44	142.88	44.40	192.28	53.93

Between groups analysis was conducted using a one-way ANOVA at each time for each measure. Levene's test for all ANOVAs were non-significant, indicating equal variances between groups. There were no statistical differences between groups on any of the measures of complexity or accuracy. However, there were statistical differences between groups on W/15M, $F(2, 40) = 24.075$, $p < .001$. Post hoc tests revealed there were significant differences at Time 1, $F(2, 40) = 7.765$, $p < .001$. Therefore, an analysis of covariance (ANCOVA) was used with Time 1 as the covariate. The results of the ANCOVA at Time 2 revealed that there were statistically significant differences between groups, $F(2, 39) = 8.305$, $p < .001$. Post hoc tests revealed significant differences between the timed-writing and translation ($d = 1.49$) and sentence-combining ($d = 2.08$) groups. There were also significant differences between groups at Time 3, $F(2, 39) = 15.329$, $p < .001$. Post hoc tests at Time 3 revealed significant differences between all three groups with large effect sizes between translation and sentence-combining ($d = 1.06$) and timed-writing groups ($d = 1.70$), and between sentence-combining and timed-writing groups ($d = 2.50$).

Qualitative Data

A qualitative data component was used to examine learner perspectives on their own writing practice and development as well as to help explain some of the results of the quantitative data. Therefore, two participants were chosen from each of the writing practice groups to be interviewed. The participants were chosen by recommendations from the teacher of each group with priority placed on choosing students who were regularly attending and completing the practice tasks diligently and conscientiously. However, only one of the two participants recommended from the sentence-combining group agreed to participate, bringing the total number of participants to five. All participants read and signed consent forms and were paid 1000 yen (approximately US $10) for their assistance.

The second author conducted semi-structured interviews with each participant on Zoom. All the interviews were conducted online in their first language for approximately an hour each time. Interviews with each participant took place on two separate occasions. The first interviews were conducted during week 6 (Interview 1) and the second interviews took place during week 15 (Interview 2). These two periods of time were chosen to document changes the participants experienced with the tasks as well as their perceptions of the tasks.

With the participants' agreement, the interviews were video recorded using Zoom. The second author analyzed the interview data. First, the researcher listened to the interview

data over and over again and transcribed excerpts that were considered important. Then, the researcher examined cases both individually and then collectively to understand how they were similar and different. Factors which might have led to the variance in the data and the reasons for the variations were a particular focus. An inductive approach was used, which involved analyzing the data by focusing on specific pieces of information and by making connections among them for meaningful patterns (Hatch, 2002).

Four themes were identified, including teacher feedback, the transfer of the practice to writing assignments, the importance of including a variety of other types of practice, and aspects of the practice participants paid attention to. Regarding teacher feedback, participants expressed their desire to have teacher feedback on the practice tasks as reflected by comments such as "I need someone to judge if my sentences are correct or not" (SC student, Interview 1). Teacher feedback on the treatment materials was not permitted during the study because it would have been a confounding variable that potentially could have influenced the results of the study. Although model answers were provided, participants in all three groups felt it would have been more beneficial to get learner specific feedback to help them understand more about why they had made mistakes and more importantly how to improve their writing.

A second theme that emerged from the data was the idea of transferring the practiced skills to other skill areas and vice versa. For example, one student in the translation group discussed how the translation tasks helped her become more comfortable expressing her own ideas in English. She explained that she used to rely on a software translation site to prepare her script for presentations in other classes. However, since she learned how to use simple sentences through the task, she only used the site to look up words she did not know and instead created her own sentences by using simple forms. She said, "I feel using simple sentences makes it easier for me to speak" (TR Student 2, Interview 2). Furthermore, the participant from the sentence-combining group discussed how practicing this task helped him analyze and understand sentence patterns in English reading for class and standardized tests like TOEIC.

A third theme that emerged was the importance of other types of practice. For example, one of the participants from the translation group mentioned that she felt the translation tasks were not enough to help her improve and that learning how to compose more paragraphs in class was helpful. The participant from the sentence-combining group mentioned the importance of fluency writing. He said, "(SC) is for learning how to write basic sentences, and in terms of learning different ways to express ideas, I feel it is a bit insufficient by itself" (SC student, Interview 2). Finally, the participants in the timed-writing group did not feel they had enough practice in composing more sophisticated sentences as reflected in comments such as "[Sentence structures she writes in her timed writing are] too easy…even when I feel I can connect this part and that part, I end it with a period, so I would like to do that kind of practice [sentence-combining]" (TW Student 2, Interview 1)

Finally, the theme of attention in writing emerged from the data. Attention in this case refers to what participants reported they focused on during the practice tasks. For example, participants in the translation group discussed the importance of understanding the meaning and context of each sentence, either through understanding the original Japanese meaning first or by understanding the sentences that came before or after the target sentence. Meanwhile, the timed-writing participants discussed their focus on word counts and trying to increase the number of words for each timed-writing task. For example, one of the students in the timed-writing group said, "I try not to use new things [structure and vocabulary], and that's because I am rushing when I write. I feel I prioritize word count" (TW Student 1, Interview 1).

Discussion

The purpose of this exploratory study was to examine the effect of three different types of writing practice on the development of complexity, accuracy, and fluency in L2 writing. Before discussing the results, it should be noted that due to the small n-size and unbalanced groups, results of the statistical analyses should be interpreted with caution.

In regard to complexity, the hypothesis that practicing sentence-combining would aid the development of complexity was partially supported. Although the sentence-combining group did not show statistically significant differences over the other groups in this study, the group did show more growth and larger effect sizes than the other groups on all of the complexity measures across time. The effect was most pronounced on MLS, a global measure of sentence length. The combination of increases of C/T and T/S indicate that the growth in sentence length was influenced by the use of more subordination and coordination in this group. This finding is in line with the results reported by Marlowe (2019), in which the sentence-combining group in that study also used more coordination and subordination over time. It was also supported by the qualitative data in which the participant from the sentence-combining group noted how "it is important to understand how conjunctions are used…how adverbs such as 'in addition' are used…I would like to use it more in my writing" (SC Student, Interview 2). Not only did this participant acknowledge the importance of these structures, but also indicated a desire to integrate them more into English writing compositions.

In terms of accuracy, the means of all three groups declined across time. Only the translation group showed significant decreases across time. This result did not support the hypothesis that practicing translation would lead to increases in accuracy. Although the decline in accuracy was not statistically significant for the other groups, the overall decline across all three groups could potentially indicate trade-offs (Skehan, 2009). Growth in either complexity (SC & TR) or in fluency (all three groups) could have come at the expense of accuracy. Essentially, as learners either expand their repertoire of sentence structures or increase the amount of production, it reduces the overall accuracy. This finding is in line with the pattern of development described by Housen et al. (2012), where periods of growth in complexity were accompanied by low accuracy. Moreover, growth in fluency can often result in diminished accuracy (Skehan, 2009). Trade-offs for the timed-writing group also were indicated by comments made in participant interviews, with one participant admitting "I am trying to get more word counts…I think I make more grammar mistakes [compared to before]" (TW Student 1, Interview 2) and the other participant conceding, "I am not sure and worried if my grammar is correct" (TW Student 2, Interview 1).

Finally, for fluency, all groups made gains over the course of the 15 weeks, but the translation and timed-writing groups made statistically significant gains across time. Furthermore, the timed-writing group made gains that were statistically different from both the translation and sentence-combining groups. Therefore, the hypothesis that timed writing would have the most significant effect on fluency was supported by the results. Again, looking at the comments made by participants in the timed-writing group, there was a clear focus on prioritizing production over grammatical correctness.

Implications, Limitations, and Conclusion

This study has a few implications for L2 writing instruction and research. First, it appears that certain tasks can target developing specific components of L2 writing. For example, practicing sentence-combining hastens the development of complexity, which is in line with previous studies conducted by the main author (Marlowe, 2016, 2019). Furthermore, timed writing had a significant effect on the development of fluency. Second, it appears that growth in complexity and fluency might come at the cost of accuracy. As emergent L2 writers begin to experiment and expand their writing capabilities, they are bound to make more mistakes in the process. Finally, in regard to student perspectives, it appears that teacher feedback and guidance with writing practice tasks can help students become more comfortable and confident with L2 writing composition. Furthermore, in the interviews, the participants said that they felt that the use of different types of writing practice tasks would be more beneficial for their learning and development. Participants were restricted to one specific type of writing practice throughout the duration of the study. Most likely, a balanced approach using all three tasks as well as other forms of writing practice would provide learners more novelty as well as the chance to develop other components of their writing proficiency.

There were several limitations to the study. As mentioned above, the group sizes were small and unbalanced which made for less robust statistical findings. Second, the study relied on the use of intact classes instead of randomized groups which led to potential confounding variables including different instructors and different levels of proficiency between classes. Furthermore, participants for the qualitative component were not necessarily representative of all participants in the study. Third, time on task was slightly different for the timed-writing group, although the amount of writing between all groups was roughly similar. Finally, the study was conducted over 15 weeks, which is probably too short of a time span to capture changes in writing development that occur over longer periods of time. The research was intended to be a one-year longitudinal study, but due to the onset of the COVID-19 pandemic, the researchers were unable to gather data during the first semester of the academic year due to the sudden shift to emergency remote teaching by the institution where the study took place. Future researchers in this area could explore adding lexis as a construct, examine the effects of other writing practice tasks, and conduct the study over longer periods of time with larger groups and different types of learners.

Acknowledgements

This work was supported by Japan Society for the Promotion of Science KAKENHI (grant number JP 20K00823)

Bio Data

J. Paul Marlowe is an associate lecturer of English at Kwansei Gakuin University's School of Policy Studies. He earned a PhD in Education from Temple University Japan and is interested in researching L2 writing development. <itz97680@kwansei.ac.jp>

Mayumi Asaba is an associate professor at Kyoto Sangyo University's Faculty of Foreign Studies. She earned a PhD in Education from Temple University Japan and is interested in researching expertise in L2 teaching. <masaba@cc.kyoto-su.ac.jp>

References

Baba, K. (2020). Exploring dynamic developmental trajectories of writing fluency. Complex Dynamic Systems Theory and L2 Writing Development. In G. G. Fogal, M. H. Verspoor (Eds.), Complex *Dynamic Systems Theory and L2 Writing Development* (pp. 3-25). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.54

Baba, K., & Nitta, R. (2014). *Phase transitions in development of writing fluency from a complex dynamic systems perspective*. Language Learning, *64*(1), 1-35. https://doi.org/10.1111/lang.12033

Byrnes, H., Maxim, H. H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment. *The Modern Language Journal*, i-235. Retrieved from http://www.jstor.org/stable/40985261

Cooper, C. R. (1973). An outline for writing sentence-combining problems. *English Journal*, *62*(1), 96-108. doi:10.2307/814087

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Sage .

Ellis, R. (1994). *The study of second language acquisition* (2nd ed.). Oxford University Press.

Hatch, J. A. (2002). *Doing qualitative research in education settings*. State University of New York Press.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*, 461-473. doi:10.1093/applin/amp048

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. doi:10.1075/lllt.32.01ho

Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (Research Report No. 3). National Council of Teachers of English.

Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, *4*(1), 51-69. doi:10.1016/1060-3743(95)90023-3

Källkvist, M. (2008). L1-L2 translation versus no translation: A longitudinal study of focus- on-formS within a meaning-focused curriculum. In L. Ortega & H. Byrnes (Eds.), *The Longitudinal Study of Advanced L2 Capacities*, (pp. 182–202). Routledge.

Kuiken, F., Vedder, I., & Gilabert, R. (2010). Communicative adequacy and linguistic complexity in L2 writing. In I. Bartning, M. Martin, I. Vedder, G. Pallotti (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (EUROSLA monographs series, 1), European Second Language Association, [S.l.] (2010), pp. 81-99.

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

Lawlor, J. (1983). Sentence combining: A sequence for instruction. *The Elementary School Journal*, *84*(1), 53-62. Retrieved from http://www.jstor.org/stable/1001129

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474-496. doi:10.1075/ijcl.15.4.02lu

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1) 36-62. Retrieved from http://www.jstor.org/stable/41307615

Marlowe, J. P. (2016). Developing syntactic complexity in L2 writing. In P. Clements, A. Krause, & H. Brown (Eds.), *Focus on the learner* (pp. 397-404). Tokyo, Japan: JALT.

Marlowe, J. P. (2019). *The Effects of Sentence-Combining on The Longitudinal Development of Syntactic Complexity in L2 Writing*. Temple University.

Marlowe, J. P., & Asaba, M. (2017). Investigating the cognitive processes of translation writing tasks. In P. Clements, A. Krause, & H. Brown (Eds.), *Transformation in language education*. (pp. 369-376). Tokyo: JALT.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555-578. doi:10.1111/0023-8333.00136

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*(4), 492-518. doi:10.1093/applin/24.4.492

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

Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language learning*, *47*(1), 101-143. https://doi.org/10.1111/0023-8333.31997003

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching, 43*(1), 1-32. https://doi.org/10.1515/iral.2005.43.1.1

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510-532. doi:10.1093/applin/amp047

Skehan, P. (2018). *Second language task-based performance: Theory, research, assessment*. Routledge.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics: Pearson new international edition*. Pearson.

Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing*, *26*(3), 445-466. https://doi.org/10.1177/0265532209104670

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity* (No. 17). University of Hawaii Press.

Yoon, H. J. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System*, *66*, 130-141. https://doi.org/10.1016/j.system.2017.03.007

Appendix

Complete Descriptive Statistics for Each Measure, Group, and Time

Table A1

Descriptive Statistics for Mean Length of Sentence for Each Group Across Three Times

	Translation Group			Sentence-combining Group			Timed-writing Group		
	Time 1	**Time 2**	**Time 3**	**Time 1**	**Time 2**	**Time 3**	**Time 1**	**Time 2**	**Time 3**
M	9.06	10.34	10.05	8.12	10.86	11.02	10.21	11.94	11.75
SE	.66	.56	.70	.52	.81	.77	.78	.87	.88
95% CI	[7.66, 10.46]	[9.16, 11.53]	[8.57, 11.54]	[7.02, 9.23]	[9.16, 12.56]	[9.40, 12.63]	[8.41, 12.02]	[9.93, 13.94]	[9.72, 13.77]
SD	2.62	2.22	2.79	2.22	3.42	3.25	2.35	2.60	2.64
Skewness	.50	1.12	2.55	2.11	1.06	.38	1.16	.48	.24
SES	.56	.56	.56	.54	.54	.54	.72	.72	.72
Kurtosis	-.18	1.33	8.27	6.75	2.30	.81	.82	-.42	-.77
SEK	1.09	1.09	1.09	1.04	1.04	1.04	1.40	1.40	1.40

Note. 95% CI = 95% confidence interval; M = mean; SE = standard error; SD = standard deviation; SES = standard error skewness; SEK = standard error kurtosis.

Table A2

Descriptive Statistics for Clauses per T-Unit for Each Group Across Three Times

	Translation Group			Sentence-combining Group			Timed-writing Group		
	Time 1	**Time 2**	**Time 3**	**Time 1**	**Time 2**	**Time 3**	**Time 1**	**Time 2**	**Time 3**
M	1.22	1.37	1.32	1.19	1.38	1.34	1.40	1.51	1.44
SE	.05	.05	.04	.04	.07	.03	.08	.07	.07
95% CI	[1.12, 1.32]	[1.27, 1.46]	[1.24, 1.40]	[1.10, 1.28]	[1.24, 1.52]	[1.27, 1.41]	[1.20, 1.59]	[1.36, 1.67]	[1.28, 1.60]
SD	.18	.18	.15	.18	.28	.14	.25	.20	.20
Skewness	.47	-.03	.31	2.27	1.65	-.75	1.07	.93	.31
SES	.56	.56	.56	.54	.54	.54	.72	.72	.72
Kurtosis	-1.11	-1.26	2.00	7.59	4.27	-.40	.28	.91	-1.21
SEK	1.09	1.09	1.09	.54	.54	.54	1.40	1.40	1.40

Note. 95% CI = 95% confidence interval; M = mean; SE = standard error; SD = standard deviation; SES = standard error skewness; SEK = standard error kurtosis.

Table A3

Descriptive Statistics for T-Units per Sentence for Each Group Across Three Times

	Translation Group			Sentence-combining Group			Timed-writing Group		
	Time 1	**Time 2**	**Time 3**	**Time 1**	**Time 2**	**Time 3**	**Time 1**	**Time 2**	**Time 3**
M	1.05	1.11	1.10	1.03	1.14	1.16	1.10	1.12	1.14
SE	.03	.03	.02	.02	.03	.03	.02	.03	.03
95% CI	[.98, 1.12]	[1.05, 1.17]	[1.06, 1.14]	[.99, 1.06]	[1.08, 1.19]	[1.09, 1.23]	[1.06, 1.14]	[1.06, 1.18]	[1.06, 1.22]
SD	.13	.11	.07	.07	.12	.14	.05	.08	.10
Skewness	1.80	.39	-.26	.19	1.37	1.59	.51	.18	-.07
SES	.56	.56	.56	.54	.54	.54	.72	.72	.72
Kurtosis	3.16	-.85	-.46	.63	1.85	3.92	-1.78	-1.19	-1.58
SEK	1.09	1.09	1.09	1.04	1.04	1.04	1.40	1.40	1.40

Note. 95% CI = 95% confidence interval; M = mean; SE = standard error; SD = standard deviation; SES = standard error skewness; SEK = standard error kurtosis.

Table A4

Descriptive Statistics for Error-free Clause Ratio for Each Group Across Three Times

	Translation Group			Sentence-combining Group			Timed-writing Group		
	Time 1	Time 2	Time 3	Time 1	Time 2	Time 3	Time 1	Time 2	Time 3
M	.69	.61	.54	.66	.61	.59	.65	.64	.58
SE	.05	.04	.04	.03	.04	.04	.39	.05	.04
95% CI	[.59, .79]	[.52, .70]	[.45, .63]	[.59, .73]	[.53, .70]	[.51, .68]	[.56, .74]	[.53, .74]	[.50, .67]
SD	.19	.18	.17	.14	.17	.17	.12	.14	.11
Skewness	-.62	.04	-.26	-.30	-.23	-.35	-.41	-.24	-1.24
SES	.56	.56	.56	.54	.54	.54	.72	.72	.72
Kurtosis	.49	-1.09	.69	.14	-.75	.24	-.96	1.07	.30
SEK	1.09	1.09	1.09	1.04	1.04	1.04	1.40	1.40	1.40

Note. 95% CI = 95% confidence interval; M = mean; SE = standard error; SD = standard deviation; SES = standard error skewness; SEK = standard error kurtosis.

Table A5

Descriptive Statistics for Words per 15 minutes for Each Group Across Three Times

	Translation Group			Sentence-combining Group			Timed-writing Group		
	Time 1	Time 2	Time 3	Time 1	Time 2	Time 3	Time 1	Time 2	Time 3
M	71.47	85.00	115.97	72.92	69.75	81.50	116.44	142.88	192.28
SE	7.32	8.01	8.24	6.51	5.26	7.55	11.81	14.80	17.98
95% CI	[55.87, 87.07]	[67.92, 102.08]	[98.40, 133.54]	[59.19, 86.65]	[58.64, 80.85]	[65.57, 97.43]	[89.20, 143.69]	[108.76, 77.01]	[150.82, 33.73]
SD	29.28	32.06	32.97	27.61	22.34	32.03	65.34	35.44	44.40
Skewness	.48	.69	.30	-.17	-.35	.74	-.56	-.09	-.85
SES	.56	.56	.56	.54	.54	.54	.72	.72	.72
Kurtosis	-.73	.12	-1.04	-.73	-.74	-.79	-1.25	-2.20	.12
SEK	1.09	1.09	1.09	1.04	1.04	1.04	1.40	1.40	1.40

Note. 95% CI = 95% confidence interval; M = mean; SE = standard error; SD = standard deviation; SES = standard error skewness; SEK = standard error kurtosis.