ARTICLE

ICALL offering individually adaptive input: Effects of complex input on L2 development

Xiaobin Chen, University of Tübingen

Detmar Meurers, University of Tübingen

Patrick Rebuschat, Lancaster University

Abstract

The Artificial Intelligence methods employed in Intelligent Computer Assisted Language Learning (ICALL) in principle makes it possible to individually support language learners. Second Language Acquisition (SLA) research and language teaching practitioners agree on the relevance of target language input adapted to the learner level. However, little systematic research has explored individually adapting input and how it impacts learners. Building on previous findings on apparent alignment between the complexity of learner input and their output (Chen & Meurers, 2019), the purpose of this study is to investigate how different challenge levels of adaptive input impact learners' written output . We developed an ICALL system implementing a Complex Input Primed Writing task that selects texts for individual learners and ran an experiment grouping learners into four classes: no, low, medium, or high challenge in relation to the individual learners' writing complexity. The results show that learners generally were able to align to low- and medium-level challenges, producing more complex writings after receiving the adaptively challenging input, but less so for the high challenge group. The study demonstrates how an ICALL system used in a regular language learning context can support SLA research into adaptive input and complexity alignment.

Keywords: *Individually Adaptive Input, Complex Input Primed Writing, Intelligent Computer Assisted Language Learning, Input Output Alignment*

Language(s) Learned in This Study: *English*

Introduction

Providing learners with individual support and adaptive input to foster their language development is a central goal of Intelligent Computer Assisted Language Learning (ICALL; Meurers, 2020). Independent of their theoretical orientation, Second Language Acquisition (SLA) researchers have long recognized the importance of providing second language (L2) learners with input that is appropriate to the learner's current proficiency level (cf., e.g., the contributions in VanPatten & Williams, 2015). While in a traditional classroom context, it is not feasible for a teacher to provide each learner with the adaptive input or individual feedback that they need, recent research illustrates that ICALL systems can provide effective individual learning support in such school contexts (Chen & Meurers, 2019; Choi, 2016; Heift, 2019; Meurers et al., 2019).

Adaptive learning environments can be implemented in a variety of ways, integrating the analysis of learning input and learner output (Heift & Schulze, 2007; Lu, 2018; Meurers, 2020) and characterizations of individual learner differences (proficiency, interests, motivation, cognitive abilities; Chapelle & Heift, 2009; Heift, 2008). In this study, we focus on adaptive input selection and its effects on L2 learning through ICALL. We demonstrate how an ICALL system can analyze *linguistic complexity*, a key construct in

contemporary SLA research (e.g., Bulté & Housen, 2012, and references therein), to estimate the learner's current developmental level and use it as the basis for selecting input that will promote further L2 development. Importantly, we also report the results of an experimental study designed to empirically assess the effectiveness of the approach in promoting the acquisition of L2 English—in line with recent calls for using ICALL systems to investigate concepts grounded in SLA theory as part of empirical studies in authentic learning environments (Meurers et al., 2019).

In what follows, we will first review previous research on the development of L2 complexity and how it can be linked to the complexity of learning input to justify the need of an intervention study to better promote complexity development. We then lay out the design of an experiment that implemented an ICALL system to provide learners with individually adaptive input by unifying the learning input and learner production spaces with automatic analysis of their complexity. We report how the presentation of input with different levels of complexity, or *learning challenges*, relates to the learners' development of production complexity and demonstrate that it is effective to use ICALL systems to promote L2 writing complexity.

Complexity in Second Language Acquisition

The constructs of linguistic complexity in general and syntactic complexity in particular have been widely used in SLA research to (a) gauge language proficiency, (b) assess production quality, and (c) benchmark language development (Chen & Meurers, 2017; Lu & Ai, 2015; Ortega, 2012). They are either used as *independent* variables to predict text readability (Collins-Thompson, 2014; Vajjala & Meurers, 2012), evaluate writing quality (Ferris, 1994; Taguchi et al., 2013) and the like, or as *dependent* variables to investigate the effects of different learning tasks on language productions (Amiryousefi, 2016; Ellis & Yuan, 2004; Ong & Zhang, 2010; Révész et al., 2017; Sotillo, 2000, 2016) as well as to characterise writings by learners of different developmental stages (Bulté & Housen, 2018; Vyatkina, 2012, 2013) and/or with different backgrounds (Lu & Ai, 2015). Complexity has proved to be an effective construct for this research. However, most of the previous studies analyzed the linguistic complexity of either learning input or learner productions *separately*. Although there is already evidence showing that the input and production spaces can be linked with linguistic complexity (Chen & Meurers, 2019)[1], there has never been intervention studies exploring the effects of complex input on learner language development. The current study tries to close this gap with an experiment to examine the effects of complex input on the longitudinal development of writing complexity.

Although the goal of L2 acquisition is not to produce complex language as such, as their proficiency increases, learners usually demonstrate better mastery and more frequent use of complex language because of their expanding linguistic repertoire and capacity to use a wider range of the linguistic resources offered by the L2's grammar (Ortega, 2015). Foster and Skehan (1996) also characterized language development as "progressively more elaborate language" and "greater variety of syntactic patterning" (p. 303). Thus it is justifiable to use complexity to gauge the development of the learners' L2, or rather, as a proxy to their L2 proficiency. It should be acknowledged that not all complexity measures are correlated with proficiency, but a number of measures (e.g., Mean Length of T-unit, or MLTU, as used in the study) have been found to be distinguishing of it (Bulté & Housen, 2018; Ortega, 2003). If we consider mastery of more complex language, which is manifested as the ability to understand and produce it appropriately, as L2 development, the question is then how this development can be better promoted.

It is widely accepted in SLA research that input is essential for L2 development to take place and that the input should be at an appropriate level. For example, if the input clearly exceeds the L2 ability of the learner, acquisition might not take place. At the same time, there is still debate on how to define *appropriate* input in this context. A classic position such as that of Krashen (1985) holds that L2 input should be sufficient, comprehensible, and at the level of *i+1*, that is, slightly above the current level of the learner's interlanguage. From a complexity point of view, the complexity of the target language input should then be slightly higher than what the learner can understand or produce—the current developmental stage of the learner. As discussed previously, linguistic complexity can be used as a proxy for development/proficiency.

It can also be used to assess the appropriateness of the input in terms of ease of understanding for learners of certain proficiency levels (e.g., to assess the readability of reading texts). However, it is still unclear whether the input chosen with the complexity analysis approach can practically promote the development of the learner's L2 and if it does, what the optimal amount of complexity difference between the input and the output should be. In other words, how big should the "+1" difference be? These are the empirical questions the current study tries to answer.

An advantage of using the complexity construct to investigate the effects of input on L2 development is that complexity analysis is automatizable. Several computer systems making use of Natural Language Processing (NLP) technologies have been developed for the automatic analysis of linguistic complexity (e.g., Chen & Meurers, 2016; Crossley et al., 2016a, 2016b; Lu, 2010, 2012; McNamara et al., 2014). They have been proven useful for research questions related to language learning (e.g., Alexopoulou et al., 2017; Kyle, 2016; Lu & Ai, 2015). However, to the best of our knowledge, no research has been done on how the automatic analysis of linguistic complexity can be combined with L2 learning theories to develop practical ICALL systems for L2 learning. The experimental system used in the current study is a prototype of such an ICALL system. As a result, the study also demonstrates how the integration of NLP technologies and SLA theories can help create a new agenda for L2 learning research.

Approach and Research Questions

Preliminary research by Chen and Meurers (2019) has shown promising results of the effects of complex input on the complexity improvement of the learners' L2 production. They used a continuation writing corpus collected by Wang and Wang (2015) who asked a group of Chinese learners of L2 English to read two stories (one in Chinese and the other in English) with endings removed and to continue writing the stories in their L2. Wang and Wang postulate that the coupling of comprehension with production in continuation writing tasks creates forced alignment between the writing and the reading, which is realized by structural priming. Priming is significant for L2 development because it increases the likelihood of the learners using structures that they come across but have not yet fully acquired through meaning-focused activities (McDonough & Mackey, 2006). Although previous research on syntactic priming mainly focuses on oral interactions, theoretically it is reasonable to believe that it can also happen in reading and writing scenarios (see Wang & Wang for detailed theoretical reasoning drawing on the Interactive Alignment Model and the Situation Model). The effect of priming lies in that it helps strengthening knowledge representation and proceduralizing linguistic forms (de Bot, 1996; Nobuyoshi & Ellis, 1993), which are both essential to SLA.

Chen and Meurers (2019) calculated two indexes from the complexity of the input stories and the students' writings:

>   *challenge = complexity (English input text) – complexity (student's baseline writing)*

and

>   *improvement = complexity (continuation writing) – complexity (student's baseline writing)*

Baseline writing refers to the English texts that the learners produced after reading the story in Chinese, their L1. Continuation writing refers to the continued English writing after learners read the story in English, their L2. High Pearson's correlation coefficients (ranging from .28 to .96, all of which significant at $p < .01$) were observed between challenge and improvement for most of the complexity measures they used. Chen and Meurers's (2019) study demonstrates that it is viable to relate learning input to learner production with complexity measures. It provides an operationalizable implementation of individually adaptive input. However, because of the limitation of the dataset, it is still unclear whether different levels of challenge would result in different learning effects and how to account for the gap between the receptive and active knowledge of the L2 (understanding and producing the language).

This leads to the purposes of the current study, which boil down to the following research questions:

1. Does the complexity of written output align to the complexity of the input?
2. If so, is there incremental alignment when the learner is presented with a sequence of increasingly complex input?
3. If so, does this incremental alignment result in lasting L2 learning?

Answers to these questions will not only provide language teachers, textbook authors, and other practitioners with concrete guidance on how to select learning input based on the evaluation of the learners' current proficiency levels, but also provide insights into the design of ICALL systems for L2 acquisition because all analysis proposed in this study are automatizable. We tried to answer the research questions with a fully automatic intervention study, whose design and procedure are introduced in the next section.

Methods

There is a general lack of intervention studies on the effects of complex input on proficiency development. Such a study would require assessment of the learners' current proficiency and assignment of input of various challenge levels. It also needs to single out a test measure (independent variable) and control for all possible confounding factors. The experimental task should not only ensure that the treatment is received by the learners but also be able to elicit production from them for the purpose of evaluating the effects of the treatment. Based on these considerations, a Complex Input Primed Writing (CIPW) task was conceived for the purpose of the current study. It is based on the continuation writing task designed by Wang and Wang (2015) but adds some modules to automatically analyze the learners' L2 production and choose reading input based on the analysis.

Automatic Analysis of Linguistic Complexity

The advantage of using linguistic complexity to assess L2 proficiency to locate input appropriate for promoting language development is that the whole process is automatizable, making it possible to provide learners with individualized and adaptive learning materials. A number of general-purpose systems have been developed for extracting complexity measures from both learning input and learner productions (e.g., Chen & Meurers, 2016; Kyle & Crossley, 2015; Lu, 2010, 2012; McNamara et al., 2014). The experiment system used in the current study is built on the basis of the Common Text Analysis Platform (CTAP; Chen & Meurers, 2016), which is capable of extracting large numbers of complexity measures from multiple lexical, syntactic, and discoursal levels. The complexity of the texts in the reading corpus was extracted and stored in a database beforehand, while the analysis of the participants' writings was done online to dynamically choose input texts for the next CIPW task.

The CIPW Tasks

The CIPW task is a task in which the participants are asked to complete a text whose ending has been removed. The genre of the text can vary, for example, it can be a narrative story or a piece of argumentative writing. We removed the last quarter of the sentences in each essay for this experiment, leaving the first 75% of the essay to be read and continued by the participants. The participants were instructed to continue writing the essay in a way that completed the narration or argumentation as coherently as possible. It is a focus-on-meaning task from the participants' perspective but offers linguistic priming (Wang & Wang, 2015) for the writing because the participants need to read and understand the remaining part of the essay in order to be able to continue writing it. In the current study, all essays used in the CIPW tasks are chosen based on two individualized criteria: (a) the baseline complexity of the previous writings by the participant and (b) the treatment group the participant is in. In order to answer whether different levels of complexity challenge would result in different improvement, four treatment groups were used: zero-, low-, medium-, and high-challenge groups.

We adopted the syntactic complexity measure of MLTU as the treatment measure of the current study. A T-unit is a minimally *terminable* unit (hence the name T-unit) or the shortest grammatically allowable sentence which consists of "one main clause with all subordinate clauses attached to it" (Hunt, 1965, p. 20).

The MLTU measure has been consistently found to discriminate L2 proficiency levels and to develop in a somewhat linear manner (Bulté & Housen, 2018; Lu, 2011; Ortega, 2003). Ortega (2003) suggested that a difference of 2 words in MLTU be seen as statistically significant to differentiate two consecutive proficiency levels. As a result, the participants in the zero-, low-, medium-, and high-challenge treatment groups were assigned texts that were +0, +2, +4, and +6 words more than their baseline MLTU respectively. The baseline MLTU was calculated as the mean MLTU of all writings a participant had submitted to the system. For example, if a participant had been able to produce writings with a mean MLTU of 10 words and was randomly assigned into the low-challenge treatment group, in the next CIPW task, she would receive texts with an MLTU of 12 words (the baseline of 10 words plus a low challenge of 2 words). If she had been assigned to the medium-challenge group, the texts she would receive in the next CIPW task would have an MLTU of 14 words instead.

In order to make sure that the texts controlled for MLTU assigned to the participants were comprehensible to them, the ICALL system we designed for the experiment chose from the reading corpus only texts that met the abovementioned criteria, as well as were closest to the participants' earlier production in all other complexity dimensions—the *nearest neighbours* in the complexity vector space. The CIPW ICALL system was able to extract 576 lexical, syntactic, and cohesion complexity measures from a text[2]. Consequently, except for the treatment measure of MLTU, all the other 575 measures were first normalized and then used to calculate the Euclidean distance between the learner production and the texts in the reading corpus. The distance between two points *p* and *q* in a Euclidean *n*-space can be calculated with the Pythagorean formula (Equation 1):

$$d(p, q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \tag{1}$$

For each CIPW task, the 10 texts from the reading corpus that were closest to earlier learner productions (10 nearest neighbours) were offered as choices for the participants to continue writing. It is reasonably assumed that the input texts that are close to the learners' production in terms of linguistic complexity are comprehensible by them. As a result, depending on the complexity of their writing, each learner would receive highly personalized and adaptive CIPW tasks.

The Reading Corpus

The corpus of reading essays for the CIPW tasks was collected from Newsela, an American educational website featuring articles on various contemporary topics that target students of different grades and reading abilities. The Newsela website adapts each published story into five different reading levels (ranging from the second to the twelfth grades), including the original version of the story as the "max" level. The reason for using the Newsela corpus in our study is that it offers a broad spectrum of variability in the linguistic complexity of the texts. This is important for a system that offers individualized input based on the learners' proficiency.

The Newsela website offers essays in different genres, including news, narratives, argumentations, and so on. However, for a controlled intervention experiment like the current study, it is important to control for genre because it has been found to affect the complexity of the learner's production in previous research (e.g., Beers & Nagy, 2009; Way et al., 2000; Yoon & Polio, 2017). For example, Yoon and Polio found that the complexity of learner-produced argumentatives is higher than that of narratives. As a result, the current study restricted the CIPW task genre to argumentative writings. Six-hundred-thirty-five texts were obtained from the "Opinion" and "Pro/Con" sections of the Newsela website. The MLTU of these texts ranged from 7.42 to 30.42 (*M* = 14.60, *SD* = 4.04). Table 1 summarises the profile of the Newsela corpus used in the present study.
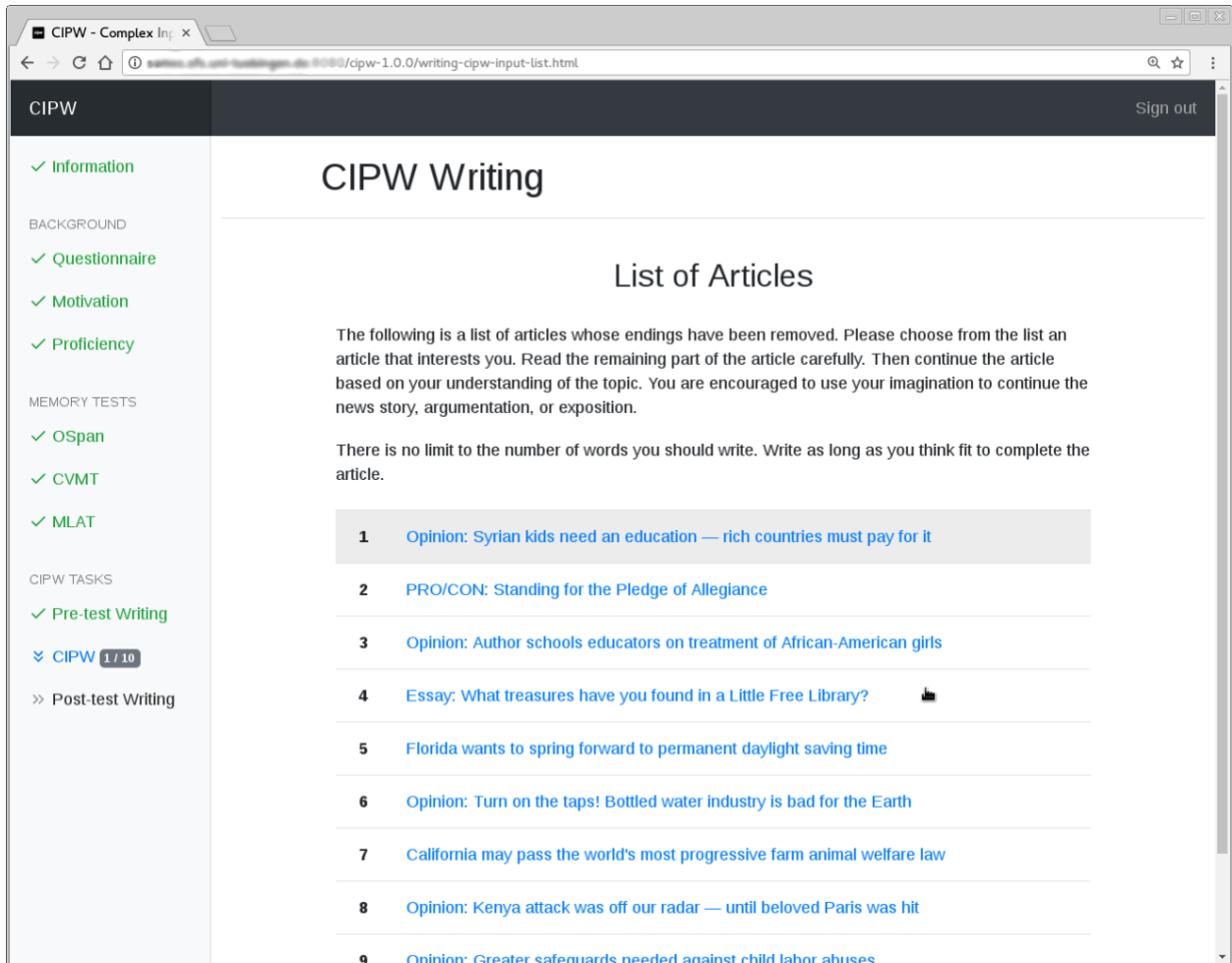
Table 1

Profile of the Reading Corpus Used in the Current Study

Grade Level	# Texts	# Words/Text	Mean MLTU	SD MLTU
2	16	427.25	8.24	0.59
3	99	502.20	9.56	0.82
4	46	704.43	11.12	0.90
5	105	804.48	12.58	1.01
6	72	956.83	14.45	1.11
7	80	993.35	15.50	1.27
8	46	1086.04	17.56	1.41
9	58	1032.05	18.33	1.74
12	113	1092.60	20.00	2.91

Procedure

The experiment was conducted in a fully automatic ICALL system we created for the study. After signing up to take part in the experiment, the participants received an email with login details to the ICALL system, which was used to collect information on the participants' background, individual difference metrics[3], proficiency test, pre- and post-test writings, as well as the intervention treatment of 10 CIPW writings. The proficiency test used in the study was a web-adapted version of the C-tests (Klein-Braley, 1985) used by the Language Learning Center at the University of Tübingen for language course placement purposes. C-tests have been found to be predictive of general L2 proficiency (Dörnyei & Katona, 1992; Eckes & Grotjahn, 2006; Harsch & Hartig, 2016). The reliability of the C-tests calculated as Cronbach's α was .93.

The difference between the pre-/post-test writings and the CIPW writings is that the former is a free-writing task with only topic prompts, while the latter is a continuation writing task which provides the participants with the first three quarters of a text with which they are required to continue writing after reading it. The pre-/post-test writing topics are "shared economy" and "work from home" respectively. The selection of the first CIPW task essays is based on the complexity of the pre-test writing. The subsequent ones are based on the mean complexity of the submitted writings of individual participants. Figure 1 shows a screenshot of the ICALL system, whose left navigation menu lists all the questionnaires and tasks the participants are required to complete in sequence. Figure 2 summarizes the procedure of the whole experiment.

Figure 1

A Screenshot of the CIPW ICALL System

Experiment Procedure

Participants

One-hundred-and-sixty-three Chinese learners of English from a Chinese university answered to the call-for-participation of the study and were sent login details to the ICALL system. Of those, 112 participants finished at least one continuation writing task and were included in the analysis. After providing informed consent, the participants were randomly assigned into a control group and the four experimental groups who received different levels of challenge, as described in the previous section. Participants in the control group did not do the CIPW writing tasks. Instead, they were required to finish the pre- and post-test writings with an interval of three weeks in between, which was also the time allowed for the other groups to finish all the writing tasks. We believed that three weeks were a period concentrated enough for us to see the effects of the treatment, if there would be any, but not too long for the treatment to be confounded with the

participants' development due to the language courses they were taking from the university. Practically, allowing too short a period for the participants to finish the CIPW tasks would risk placing too much stress on their already very busy university schedule. The pre- and post-tests were exactly the same for all the groups (control and experimental groups), but only the four experimental groups used the ICALL system to finish the CIPW tasks. We included a control group which did not use the ICALL system to control for the fact that our participants were also taking English courses from the university. The control group is thus different from the zero-challenge group. It allows us to attribute the changes between the pre- and post-tests to the use of the ICALL system rather than confounding the cause of such changes with the English courses they were taking. On average, the number of writings each participant in the experimental groups finished was 8.68 (*SD* = 4.35). Participants were rewarded with Amazon vouchers based on the number of tasks they completed (50 RMB per task). No course credits or other forms of reward were provided. The study was approved by the University Ethics Committee of the University of Tübingen.

Out of the 112 participants included in this analysis, 66 were male and 46 were female. Their ages ranged from 17 to 27 years (*M* = 18.98, *SD* = 0.88). The mean number of years they had spent learning English was 10.56 years (*SD* = 2.39). In the background questionnaire, the participants were asked to self-indicate their English proficiency with a set of proficiency descriptors. Five participants thought they were post-beginners, 23 lower-intermediates, 67 intermediates, 15 upper-intermediates, and two did not report their proficiency. No participants considered themselves beginners. The C-test results also showed that the majority of participants were intermediate learners of English. Table 2 summarizes the proficiency of the participants in each group. Although there are discrepancies between the participants' self-reported proficiency levels and the C-test results, we can still see that most of them are at the lower-intermediate or intermediate levels. As far as we know, the participants were not taking English writing courses during the process of the experiment since the university from which the participants were recruited did not offer writing courses to non-English-major students, and the participants were non-English majors except for eight pilot study participants.

Table 2

Number of Participants in Each Group and Their Proficiency Distribution Based on the C-test Results

Group	A1	A2	B1	B2	C1	C2	Total
Control	1	7	1	2	1	0	12
No-challenge	2	18	4	2	1	0	27
Low-challenge	2	12	7	2	1	1	25
Medium-challenge	0	15	4	4	1	0	24
High-challenge	1	10	5	6	2	0	24
Total	6	62	21	16	6	1	112

Results

In light of the research questions on whether complex input fosters development of L2 writing complexity and if it does, how much more complex the input should be with regard to the complexity of their writing, we first present results on the comparison of the complexity of the pre- and post-test writings. Then we explore the patterns of writing complexity across time/tasks. The interaction between the complexity of input and proficiency development is operationalized as the longitudinal interaction between the challenge the participants received and the improvement they made from each CIPW task. Detailed analysis and results are reported in the following sub-sections.

Complexity of Pre- and Post-test Writings

The variable of interest in the experiment is the MLTU of the participants' writings. Table 3 shows the MLTU of pre- and post-test writings of different experiment groups. On average, the participants were able to produce texts with an MLTU of 18.12 words (SD = 6.5) in the pre-test writing task. A one-way ANOVA confirmed that the five experiment groups did not differ in the MLTU of their initial writings: F (4, 107) = 0.11, $p \geq$ .1. Out of the 112 participants who finished the pre-test, 71 also completed the post-test. The mean MLTU of the post-test writings was 17.15 words (SD = 6.12). No significant differences were found for the post-test writing MLTU among the groups either: F (4, 66) = 0.91, $p \geq$ .1. The changes of MLTU between the pre- and post-test writings were calculated for those who finished both writings. The mean changes were negligible: M = 0.24 and SD = 6.41. Again, no significant differences were found among the experimental groups: $F(4, 66)$ = 0.43, $p \geq$ .1. Thus, the pretest–posttest analysis, prima facie, suggests that no learning took place. However, as we will show below, a closer look at the developmental pattern reveals a more elaborate picture.

Developmental Patterns of Writing Complexity

In order to observe how the complexity of the writings developed across CIPW tasks, an MLTU developmental trajectory was plotted for each individual participant[4]. The dynamically adaptive nature of the system makes it difficult to summarize the developmental trajectories, but Figure 3 shows a typical developmental trajectory for each experimental group. The plots show a wavy developmental pattern for the complexity of the writings across tasks. For most participants, the complexity of their writings increased at the beginning of the experiment before falling to the beginning level and then back up again. Although there are individual differences in the magnitude (height and width) of the "waves," the wavy pattern is observable in almost all participants who finished more than a few CIPW tasks.

Table 3

Mean and SD of MLTU of Pre- and Post-Test Writings by Experiment Group

Group	Pre-test		Post-test	
	M	*SD*	*M*	*SD*
Control	18.76	7.48	19.55	4.95
No-challenge	18.15	9.12	17.27	5.31
Low-challenge	18.56	6.91	16.00	4.29
Medium-challenge	17.54	3.96	15.54	4.08
High-challenge	17.91	4.14	18.14	11.66
All participants	18.12	6.50	17.15	6.12

Challenge and Improvement

Following Chen and Meurers (2019) and for the purpose of investigating the effect of input complexity on that of the continuation writings, the indexes of *challenge* and *improvement* were calculated (see the Approach and Research Questions section). The original setup of the ICALL system was to use the mean MLTU of all previous writings of the participant as the *baseline*. Depending on the groups the participants were assigned, they would always receive texts that were challenging with respect to this baseline. As a result, the baseline would change dynamically as the experiment progressed because it would be recalculated every time a new writing was submitted. If, for instance, a participant completed a writing with a higher MLTU than the mean MLTU of all her previous writings, the new mean MLTU would increase, resulting in a higher absolute MLTU for the next input as compared to that of the previous one, and vice

versa. The baseline calculated in this way is called the *dynamic baseline* in our analysis. Another way to calculate the baseline is to use the mean MLTU of the pre- and post-test writings, which would result in a *static baseline* for each participant, because this baseline value does not change. Since the static baseline also depicts what the participant is capable of doing in terms of writing complexity without being primed by more complex input—it is calculated with the MLTU of the pre- and post-writings—it can also be considered as the *proficiency baseline* of the learner.

Equations 2 and 3 were used to calculate the challenge and improvement indexes, where C is the complexity measure, or MLTU in the case of the experiment. C_i denotes the complexity of input for a specific CIPW task, while C_w is the complexity of the participant's writing. The baseline complexity is denoted as C_{db} or C_{sb}, for dynamic and static baselines respectively.

$$challenge = C_i - C_{db/sb} \tag{2}$$

$$improvement = C_w - C_{db/sb} \tag{3}$$

Figure 3

Example Developmental Trajectories of MLTU Across Writing Tasks by Participants of Different Experimental Groups
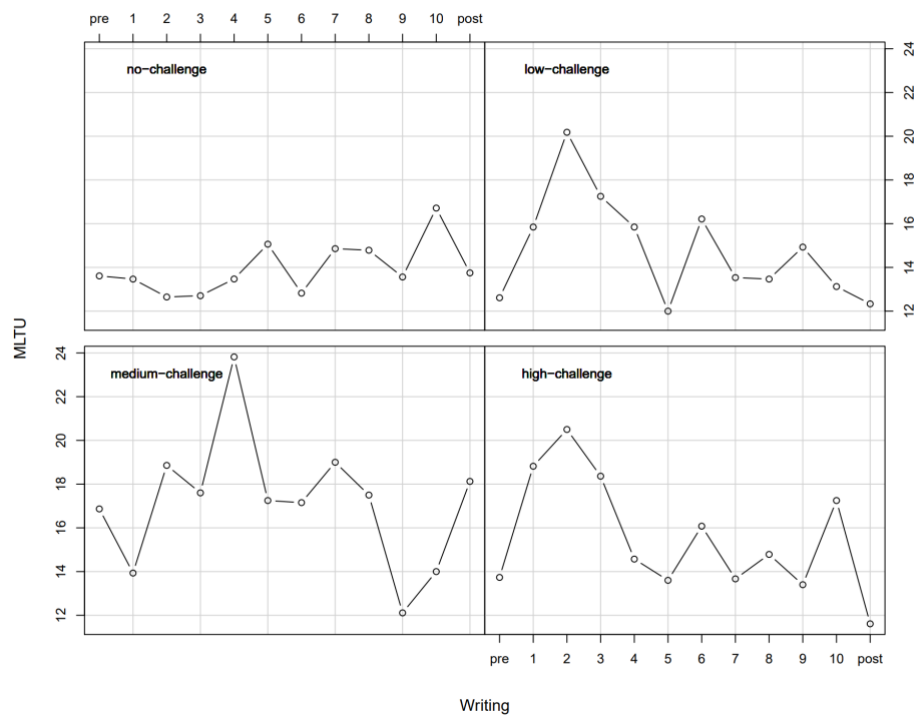
Figures 4 and 5 plot the summarized relationship between the mean challenge the participants received and the average improvement they made. Figure 4 used the dynamic baseline, while Figure 5 used the proficiency baseline to calculate the plot indexes. Linear regression models were fitted for both calculations with challenge as predictor of improvement. As is also observable from the plots, challenge does not predict improvement when the indexes are calculated with the dynamic baseline, Figure 4: $R^2 = .03$, $F(1, 89) = 2.75$, $p > .1$. In contrast, the model with indexes calculated with proficiency baseline shows a clear linear trend: challenge is highly predictive of improvement, Figure 5: $\beta = .77$, $p \leq .01$; adjusted $R^2 = .45$; $F(1, 89) = 73.59$, $p \leq .01$. Comparison of the two models shows that the static proficiency baseline better helps explain the relationship between the complexity of the input texts and that of the continuation writings.

Figure 4

Mean Improvement by Challenge with Dynamic Baseline

Challenge = input complexity − dynamic baseline
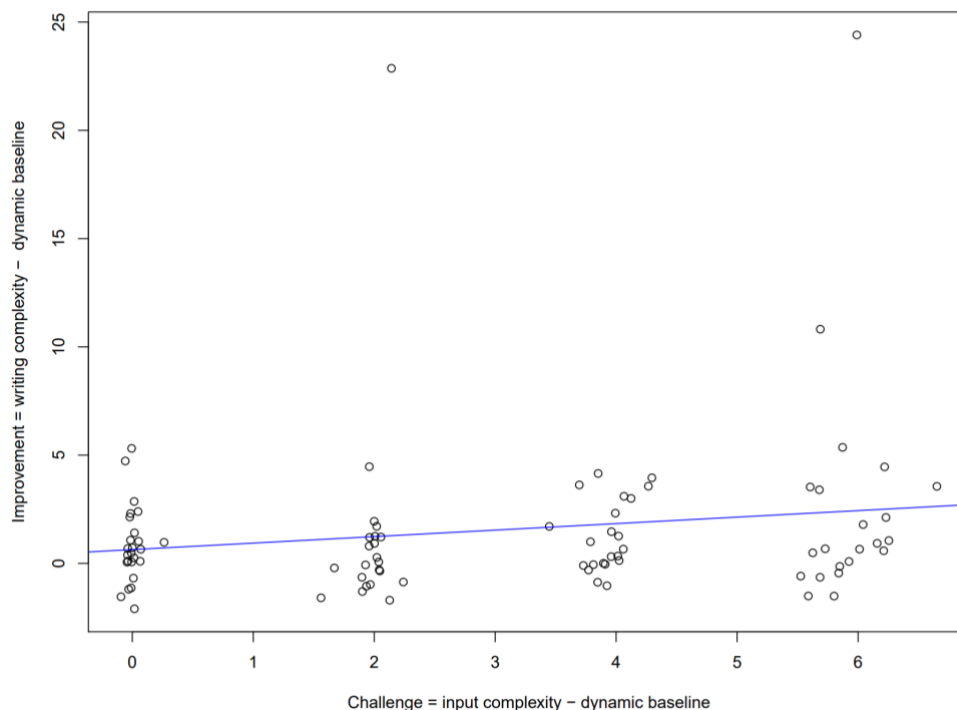
To further account for by-participant and by-task variation, a mixed-effect model was fitted with the *lme4* (Bates et al., 2015) package in R (R Core Team, 2015). The challenge calculated with proficiency baseline was entered into the model as a fixed effect of complexity improvement in the participants' writings. As random effects, participants and the sequence of completed writing tasks were entered as both random intercepts and random slopes for the effect of improvement. Equation 3 shows the configuration of the mixed-effects model used in the R environment. Residual plots of the model showed no obvious violation of homoscedasticity or normality. No interaction was found between challenge and the proficiency of the participants as assessed by the C-test. Comparison of the models with and without such interaction was done with likelihood ratio tests and yielded $x^2 = 5.32$, $p \geq .1$. In the model denoted by Equation 4, 19% of the variance in improvement was explained by the fixed terms (marginal $R^2 = .19$), while both the fixed and random factors were able to account for 72% of the same variance (conditional $R^2 = .72$).

$$imprv \sim chllng + (1 + chllng|sbjct) + (1 + chllng|wrtng) \tag{4}$$

Figure 6 shows some patterns[5] of interaction between challenge with regards to proficiency baseline and the complexity improvement of the participants' writings after receiving the challenge. Participants from the same experimental groups were plotted on the same rows. Hence, from bottom up, the rows of plot panels represent data from the no-, low-, medium- and high-challenge groups respectively. The columns of panels in Figure 6 show different interactional patterns. The left-most column, except the top panel, shows participants who were able to "outperform" the challenge, hence the black solid lines are mostly above the blue dashed ones. The second column shows participants who were able to "catch up" with the challenge, while the last two columns show participants who could barely meet the challenge or fell completely behind it. Each type of interactional pattern between challenge and improvement is observable in all experimental groups. However, the general trend is that the groups that received higher levels of challenge usually witness

more cases of failures to achieve the same levels of improvement as the challenge.

Figure 5

Mean Improvement by Challenge with Proficiency/Static Baseline
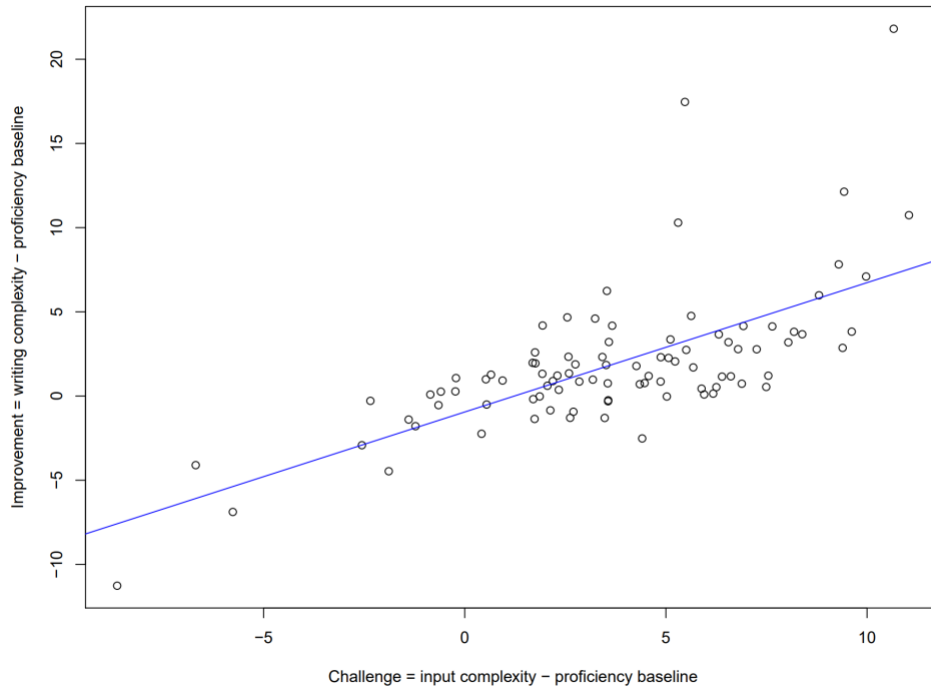
Summary of Main Results

No significant difference was found between the complexity of the pre- and post-test writings, which were two free writing tasks on different topics. However, wavy developmental patterns were observable from most participants. The complexity of the CIPW writings fluctuated in response to the complexity of the input, but within a certain limit. The response of varying writing complexity to the complexity of the input could be captured with a linear mixed effects model, which found that the complexity of the input in relation to the proficiency baseline (i.e., the challenge) was able to explain 72% of the variance in the improvement the participants made in the CIPW tasks. It was also found that, in general, the participants were able to make improvements that matched with the low- or medium-level challenge. But more students would fail to catch up if the challenge was more than two levels higher than their proficiency baseline.

Sample Patterns of Interaction Between Challenge and Improvement
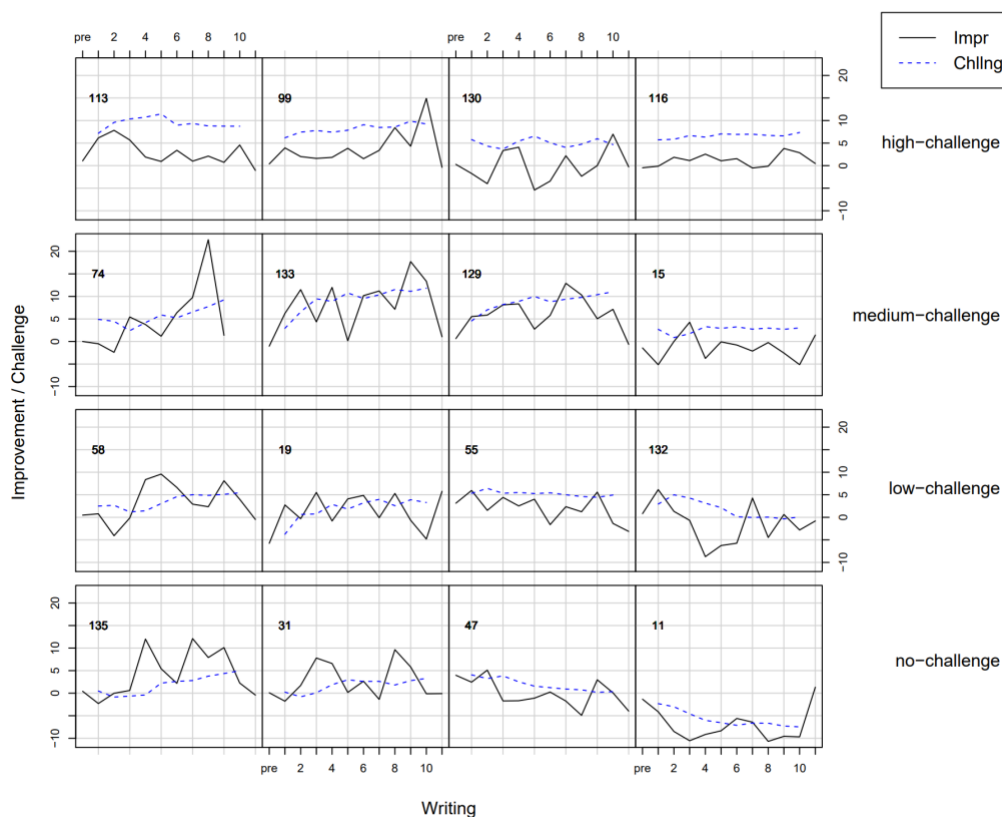
Discussion

In this study, we demonstrate how ICALL systems can provide language learners with individually adaptive input by unifying the learning input and learner production spaces with automatic analysis of their complexity. We focused on a key construct in SLA research, namely *linguistic complexity*, and evaluated empirically how the input complexity relates to production complexity.

The purpose of the current study was to investigate the effects of complex input on the development of L2 writing complexity. Specifically, we asked whether the complexity of written output would align with the complexity of learning input (RQ1), whether a sequence of increasingly complex texts could lead to incremental alignment in written output (RQ2), and whether incremental alignment could result in lasting L2 learning (RQ3). A randomized control experiment methodology was adopted with the syntactic complexity measure of MLTU as the treatment variable, while the other complexity measures were used to control for the general complexity of the input.

The developmental trajectories of the writing complexity across CIPW tasks (cf. Figures 3 and 6) indicate that the complexity of the participants' CIPW writings fluctuated in response to the complexity of the input. That is, we found evidence of alignment between the complexity of the learning input and the complexity of the written output and, more interestingly, that increasing the complexity of input texts led to a similar increment in the complexity of the output. It would be expected that the increased production complexity be carried over to free writing tasks given these observations. The fact that it did not happen may also be due to the length of the treatment. It could well be that 10 CIPW tasks are not enough to foster significant changes in proficiency, hence no significant changes in the pre- and post-test writings. Another possible

reason for this finding is task-effect, which has been found to have large effects on the complexity of L2 productions (e.g., Alexopoulou et al., 2017; Michel et al., 2007; Robinson, 2011; Tabari, 2016; Yoon & Polio, 2017; Yuan & Ellis, 2003). Even if the learners' ability to use more complex language has been expanded, when they are not primed to do so as in the situation of the free post-test writing, they would still produce language within the so-called 'comfort zone' of their abilities, rather than maximizing their complexity potentials.

The wavy developmental pattern of writing complexity for most participants reported in the previous section is not difficult to explain. Since the experiment was set up in a way that the complexity of the next CIPW input was based on the mean complexity of all previous writings of a participant, the complexity of the input would increase if the participant submitted a more complex writing. The complexity of the input kept growing if the participant was able to keep up with the challenge, until it reached a point when the participant failed to cope with the increased complexity level. The result was lower complexity of the continuation writing as compared to that of the input, drawing the average complexity of all submitted writings down if the new submission had lower complexity than the mean of all previous submissions. If the participant still failed to keep up with the new complexity level, the system would lower the input complexity again automatically until it reached a level where the participant could catch up. Then the process repeated itself, hence the multiple waves across CIPW tasks. The results show how the CIPW system is capable of dynamically adapting its tasks to the development of individual learners, which is desirable for an ICALL system providing individually adaptive input.

As for the results on the relationship between challenge and improvement, it was found that only when both indexes were calculated with the participants' proficiency baseline was challenge able to predict improvement. This suggests that the static proficiency baseline is a more accurate representation of the learners' L2 proficiency from the complexity perspective than the dynamic baseline. Combined with the finding that no difference was found between the pre- and post-test writing complexity, it can be concluded that the complexity improvement the participants gained during the CIPW tasks should not be considered as promotion of proficiency levels. At least the ability to use more advanced language had not yet been integrated as part of the learners' stable proficiency. This stabilization may require more practice. Another explanation may be that the participants had already been able to produce the more complex language but due to task or other factors, they did not use the advanced language in the free writing tasks. They would do so only when they were primed by the more challenging input in the CIPW tasks. A reviewer to an earlier version of the paper pointed out that there may be a ceiling effect because some participants were already able to produce texts with MLTU at as high a level as the highest-level texts from the reading corpus. We would like to argue that this is not the case. The most likely scenario where a ceiling effect would happen is when a participant is already producing high MLTU and is assigned to a high challenge group. In this case the system would fail to find reading texts with high enough MLTU as treatment to the participant. We receive two such reports during data collection and the experiment was stopped for these two participants. For the ones who finished all the ten CIPW tasks and the post-test writing, the system was always able to find appropriately challenging texts for the next task. In other words, if there had been a ceiling effect, the participants would not have gone through the whole test process and done the post-test.

With the proficiency baseline, even without controlling for participants and writing tasks, it was found that challenge was able to explain 19% of the variance in improvement. When they were controlled for, 72% of the variance in improvement was explainable by challenge complexity. These results suggest that although the participants were not explicitly informed about the characteristics of the input they received, they still adapted the complexity of their writings to that of the input. This phenomenon may be seen as a type of implicit learning and priming effect. Implicit learning happens when L2 learners do not pay conscious attention to meaning negotiation or sentence construction (Ellis, 2005). Structural, or syntactic, priming refers to the tendency that a speaker is more likely to use the same syntactic structures they have been exposed to in recent discourse over the alternatives (Bock, 1986). Although Wang and Wang's (2015) experiment did not focus on specific structures, the continuation writing paradigm employed in the CIPW tasks was argued to offer opportunities for learners to align their writings to the reading texts through

structural priming. Studies on L2 syntactic priming have found that L2 learners are more likely to advance to a higher stage in the developmental sequence if they are primed with developmentally more advanced forms (e.g., McDonough & Mackey, 2006, 2008; Shin & Christianson, 2012) because priming could strengthen knowledge representations (Nobuyoshi & Ellis, 1993) and make the retrieval of linguistic forms more proceduralized (de Bot, 1996). The challenging input that had a higher complexity level in the CIPW tasks can be considered as more advanced language with regards to the learners' proficiency. After being exposed to and primed by the complex input, participants receiving moderate levels of challenge were able to make improvement matching the challenge. As a result, we tend to believe that learning did occur after completing the CIPW tasks, although complexity increase is yet to be observed in free writing tasks.

The analysis of the interaction between the challenge and improvement trajectories (Figure 3) suggests that learners vary in how they react to challenge complexity. Some learners are capable of coping with both medium and low levels of challenge, while others struggle even with low challenge. However, most learners in the high-challenge group failed to catch up with the challenge. This finding confirms the importance of providing L2 learners with comprehensible input (Krashen, 1985) or input that is within their Zone of Proximal Development (ZPD, Vygotsky, 1978). The CIPW experiment makes Krashen's Input Hypothesis and Vygotsky's ZPD Theory concrete and empirically testable. The implementation of the automatic CIPW procedure makes it possible to transfer the findings in this study into practical ICALL systems.

Finally, as mentioned above, results from the experiment show that, regardless of the challenge the participants received during the 10 CIPW writing tasks, the complexity of their post-test writings did not differ significantly from the pre-test writings. That is, we found no evidence that incremental alignment could result in lasting L2 learning. This is a somewhat disappointing but also understandable result. In the current study, L2 proficiency was operationalized as the complexity of the L2 learners' writing and, to be more specific, as the syntactic complexity measured by MLTU. Although this is a very narrow and restricted view of L2 proficiency, previous research has consistently found MLTU to be the most distinguishing complexity measure of L2 proficiency (Bulté & Housen, 2018). It was also found to develop somewhat linearly across proficiency levels (Lu, 2011; Ortega, 2003). However, MLTU should not be used as a sole measure of proficiency or its development because it only accounts for one aspect of the language produced by the learner, namely the general syntactic complexity aspect. Furthermore, linguistic complexity in general and syntactic complexity in particular are easily influenced by the tasks that elicited the language production (Alexopoulou et al., 2017; Michel et al., 2007; Robinson, 2011; Yoon & Polio, 2017; Yuan & Ellis, 2003). This makes it especially problematic to use a single complexity measure as the only measure of proficiency when only one type of writing task is involved. However, for the purpose of the current study, which is about the effects of complex input on the complexity of L2 writing output, singling out the measure of MLTU allows us to investigate the interaction between input and output. As a result, caution needs to be taken when interpreting and applying the findings from this study.

The fact that the treatment of 10 CIPW tasks did not promote the development of MLTU in free writing tasks could be interpreted as no learning effect of the intervention. However, given the evidence for alignment in the developmental trajectories of the writing complexity across CIPW tasks (see above for discussion), we would argue that there is evidence of learning but that the effect might be more transitory.

Conclusion

The current study is built upon the previous finding that the spaces of L2 learning input and learner production is relatable by a common analysis of their complexity (Chen & Meurers, 2019). While most previous studies on linguistic complexity tend to characterize input or production separately, our CIPW experiment was designed to bring the two aspects of L2 learning together. The main interests of the study were on whether the complexity of learning input would affect the production complexity, which was seen as a proxy to the proficiency of the learners' L2. Results from the experiment suggest that L2 learners can be implicitly primed by syntactically more complex or more advanced language with respect to their proficiency as measured by the complexity of their L2 production. It is believed that this priming effect will

lead to L2 learning and ultimately increased L2 writing complexity, although this effect has yet to be detected in free writing tasks.

One contribution the study offers to SLA research is that it operationalized and empirically tested Krashen's (1985) widely acknowledged Input Hypothesis, or *i + 1*, whose greatest criticism is its testability and operationalizability (Ellis, 1990). The study also added a new perspective for investigating linguistic priming—from the complexity point of view rather than the traditional lexical and syntactic perspectives. Last but not least, since the accurate analysis of linguistic complexity of both learning input and learner production is highly automatizable due to the latest development of computational linguistics and natural language processing, the automatic CIPW task procedure can be integrated into practical ICALL systems to provide L2 learners with individualized and adaptive learning opportunities. The traceable developmental trajectories of the learner from such systems will also shed further light on the effectiveness and working mechanisms of complex input.

It should be acknowledged that the study also suffers from a few limitations. Firstly, a single syntactic complexity measure was used as a proxy to L2 proficiency. Although MLTU has been found to be most predictive of L2 proficiency levels, it is over-simplistic to consider it as the whole of the proficiency construct. As hundreds of complexity measures have been devised by previous research on complexity, future research should also focus on the co-varying factors instead of individual measures. Another interesting direction is to explore how the complexity factors interact with each other when they are used as complex input for L2 learning purposes, ideally also taking into account individual learner differences. Furthermore, we were unable to investigate the long-term effect of complex input on the development of general L2 proficiency due to the limited number of CIPW treatments. Future research could tackle this problem with more treatments and over a longer period of time.

Notes

1. In the study, Chen and Meurers represented learning input and learner production as multidimensional vectors of complexity measures and found that the both spaces are linkable with the Euclidean distances between the vectors. For more details on using the vector distance approach to connect input and production spaces, please refer to the original study.
2. A complete list of all the measures can be found in the supplementary materials online.
3. It should be noted that although we collected extensive data on the participants' motivation to learn a foreign language, their working memory, and declarative memory, for the purpose of the current study, these data have not been used.
4. Please refer to the supplement materials for a complete list of developmental trajectories for each participant.
5. A complete list of developmental interrelationship between challenge and improvement for all participants is included in the supplementary materials.

References

Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques. *Language Learning*, *67*(S1), 180–208. https://doi.org/10.1111/lang.12232

Amiryousefi, M. (2016). The differential effects of two types of task repetition on the complexity, accuracy, and fluency in computer-mediated L2 written production: A focus on computer anxiety. *Computer Assisted Language Learning*, *29*(5), 1052–1068. https://doi.org/10.1080/09588221.2016.1170040

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

Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing*, *22*(2), 185–200. https://doi.org/10.1007/s11145-007-9107-5

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

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

Bulté, B., & Housen, A. (2018). Syntactic complexity in L2 writing: Individual pathways and emerging group trends. *International Journal of Applied Linguistics*, *28*(1), 147–164. https://doi.org/10.1111/ijal.12196

Chapelle, C. A., & Heift, T. (2009). Individual learner differences in CALL: The FID construct. *CALICO Journal*, *26*(2), 246–266.

Chen, X., & Meurers, D. (2016). CTAP: A web-based tool supporting automatic complexity analysis. In D. Brunato, F. Dell'Orletta, G. Venturi, T. François, & P. Blache (Eds.), *Proceedings of the workshop on computational linguistics for linguistic complexity (CL4LC)* (pp. 113–119). The COLING 2016 Organizing Committee.

Chen, X., & Meurers, D. (2017). Challenging learners in their individual Zone of Proximal Development using pedagogic developmental benchmarks of syntactic complexity. In E. Volodina, G. Grigonytė, I. Pilán, K. N. Björkenstam, & L. Borin (Eds.) *Proceedings of the joint workshop on NLP for computer assisted language learning and NLP for language acquisition* (pp. 8–17). Linköping University Electronic Press.

Chen, X., & Meurers, D. (2019). Linking text readability and learner proficiency using linguistic complexity feature vector distance. *Computer Assisted Language Learning*, *32*(4), 418–447. https://doi.org/10.1080/09588221.2018.1527358

Choi, I.-C. (2016). Efficacy of an ICALL tutoring system and process-oriented corrective feedback. *Computer Assisted Language Learning*, *29*(2), 334–364. https://doi.org/10.1080/09588221.2014.960941

Collins-Thompson, K. (2014). Computational assessment of text readability: A survey of past, present, and future research. *International Journal of Applied Linguistics*, *165*(2), 97–135. https://doi.org/10.1075/itl.165.2.01col

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016a). Sentiment analysis and social cognition engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior Research Methods*, *49*(3), 803–821. https://doi.org/10.3758/s13428-016-0743-z

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016b). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, *48*(4), 1227–1237. https://doi.org/10.3758/s13428-015-0651-7

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

Dörnyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing*, *9*(2), 187–206. https://doi.org/10.1177/026553229200900206

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, *23*(3), 290–325. https://doi.org/10.1191/0265532206lt330oa

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*(2), 305–352. https://doi.org/10.1017/S027226310505014X

Ellis, R. (1990). *Instructed second language acquisition*. Blackwell.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*(1), 59–84. https://doi.org/10.1017/S0272263104026130

Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly*, *28*(2), 414–420.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*(3), 299–323.

Harsch, C., & Hartig, J. (2016). Comparing C-tests and yes/no vocabulary size tests as predictors of receptive language skills. *Language Testing*, *33*(4), 555–575. https://doi.org/10.1177/0265532215594642

Heift, T. (2008). Modeling learner variability in call. *Computer-Assisted Language Learning*, *21*(4), 305–321. https://doi.org/10.1080/09588220802343421

Heift, T. (2019). A longitudinal observation of technology-mediated feedback for L2 learners of German. *International Journal of Applied Linguistics*, *170*(2), 154–179.

Heift, T., & Schulze, M. (2007). *NLP in CALL*. Routledge.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (ED113735). ERIC. https://files.eric.ed.gov/fulltext/ED113735.pdf

Klein-Braley, C. (1985). A cloze-up on the C-Test: A study in the construct validation of authentic tests. *Language Testing*, *2*(1), 76–104. https://doi.org/10.1177/026553228500200108

Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.

Kyle, K. (2016*). Measuring syntactic development in l2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* [Unpublished doctoral dissertation]. Georgia State University.

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

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474–496. https://doi.org/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. https://doi.org/10.5054/tq.2011.240859

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, *96*(2), 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x

Lu, X. (2018). Natural language processing and intelligent computer-assisted language learning (ICALL). In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching*. Wiley-Blackwell. https://doi.org/10.1002/9781118784235.eelt0422

Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, *29*, 16–27. https://doi.org/10.1016/j.jslw.2015.06.003

McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning*, *56*(4), 693–720. https://doi.org/10.1111/j.1467-9922.2006.00393.x

McDonough, K., & Mackey, A. (2008). Syntactic priming and ESL question development. *Studies in Second Language Acquisition*, *30*(1), 31–47. https://doi.org/10.1017/S0272263108080029

McNamara, D. A., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with coh-metrix*. Cambridge University Press.

Meurers, D. (2020). Natural language processing and language learning. In C. A. Chapelle (Ed.), *The concise encyclopedia of applied linguistics* (pp. 817–831). John Wiley & Sons.

Meurers, D., De Kuthy, K., Nuxoll, F., Rudzewitz, B., & Ziai, R. (2019). Scaling up intervention studies to investigate real-life foreign language learning in school. *Annual Review of Applied Linguistics*, *39*, 161–188. https://doi.org/10.1017/S0267190519000126

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching*, *45*(3), 241–259. https://doi.org/10.1515/iral.2007.011

Nobuyoshi, J., & Ellis, R. (1993). Focused communication tasks and second language acquisition. *ELT Journal*, *47*(3), 203–210. https://doi.org/10.1093/elt/47.3.203

Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, *19*(4), 218–233. https://doi.org/10.1016/j.jslw.2010.10.003

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. https://doi.org/10.1093/applin/24.4.492

Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. In B. Szmrecsanyi & B. Kortmann (Eds.), *Linguistic complexity: Second language acquisition, indigenization, contact* (pp. 127–155). de Gruyter.

Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing*, *29*, 82–94. https://doi.org/10.1016/j.jslw.2015.06.008

R Development Core Team. (2015). *R: A language and environment for statistical computing* [Computer software manual]. The R Foundation. http://www.R-project.org/

Révész, A., Kourtali, N. E., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*, *67*(1), 208–241. https://doi.org/10.1111/lang.12205

Robinson, P. (2011). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3–37). John Benjamins.

Shin, J.-A., & Christianson, K. (2012). Structural priming and second language learning. *Language Learning*, *62*(3), 931–964. https://doi.org/10.1111/j.1467-9922.2011.00657.x

Sotillo, S. M. (2000). Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning & Technology*, *4*(1), 82–119. http://doi.org/10125/25088

Sotillo, S. M. (2016). An update on discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning & Technology*, *20*(2), 166–171. https://doi.org/10125/44470

Tabari, M. A. (2016). The effects of planning time on complexity, accuracy, fluency, and lexical variety in L2 descriptive writing. *Asian-Pacific Journal of Second and Foreign Language Education*, *1*(1). https://doi.org/10.1186/s40862-016-0015-6

Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *TESOL Quarterly*, *47*(2), 420–430. https://doi.org/10.1002/tesq.91

Vajjala, S., & Meurers, D. (2012). On improving the accuracy of readability classification using insights from second language acquisition. In *Proceedings of the 7th workshop on the innovative use of NLP for building educational applications* (pp. 163–173). Association of Computational Linguistics.

VanPatten, B., & Williams, J. (Eds.). (2015). *Theories in second language acquisition: An introduction* (2nd ed.). Routledge.

Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *The Modern Language Journal*, *96*(4), 576–598. https://doi.org/10.1111/j.1540-4781.2012.01401.x

Vyatkina, N. (2013). Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. *The Modern Language Journal*, *97*(S1), 11–30. https://doi.org/10.1111/j.1540-4781.2012.01421.x

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, C., & Wang, M. (2015). Effect of alignment on L2 written production. *Applied Linguistics*, *36*(5), 503–526. https://doi.org/10.1093/applin/amt051

Way, D. P., Joiner, E. G., & Seaman, M. A. (2000). Writing in the secondary foreign language classroom: The effects of prompts and tasks on novice learners of French. *The Modern Language Journal*, *84*(2), 171–184. https://doi.org/10.1111/0026-7902.00060

Yoon, H.-J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly*, *51*(2), 257–301. https://doi.org/10.1002/tesq.296

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*(1), 1–27. https://doi.org/10.1093/applin/24.1.1