Studies in Second Language Learning and Teaching
Department of English Studies, Faculty of Pedagogy and Fine Arts, Adam Mickiewicz University, Kalisz
SSLLT 15 (3). 2025. 501-532. Published online: 10.04.2025
https://doi.org/10.14746/ssllt.28006
http://pressto.amu.edu.pl/index.php/ssllt

The relationship between complexity, accuracy and fluency in L2 English speech: Individual differences and dynamic patterns

Cha Li
Northeast Normal University, Changchun, China
https://orcid.org/0000-0002-5819-2012
lic516@nenu.edu.cn

Mingcai Sui ✉
Changchun College of Humanities & Sciences, China
https://orcid.org/0000-0003-4989-5880
suimc335@nenu.edu.cn

Abstract

This study examined the relationship between complexity, accuracy and fluency (CAF), using longitudinal data. Conducted from the perspective of complex dynamic systems theory (CDST), this research adopted four CAF general measures to track the development of five Chinese undergraduates' English speech over a 15-week period. LOWESS graphs supported by correlation analyses showed that there were individual differences in CAF relations. Different trade-off effects were found in the CAF development of four individuals. One learner improved in lexical complexity at the expense of accuracy, another learner improved in accuracy at the cost of lexical complexity, one student made progress in syntactic complexity but sacrificed fluency, and one improved in accuracy and lexical complexity while compromising fluency and syntactic complexity, respectively. Furthermore, the use of moving correlations demonstrated that CAF relations changed dynamically over time, revealing eight identified change patterns. These patterns exhibited varying degrees of variability, ranging from highest to medium to low-medium to lowest. Six patterns involved state changes, transitioning from a competitive to supportive relationship, for example.

Keywords: complexity; accuracy; fluency; CDST; English speech

1. Introduction

Complexity, accuracy and fluency (CAF), the well-acknowledged constructs of second language performance and development (Larsen-Freeman, 2009; Michel, 2017; Norris & Ortega, 2009; Skehan, 2001), have aroused great interest in the field of second language acquisition/development (Awwad & Tavakoli, 2022; Larsen-Freeman, 2006; Li & Zhang, 2023; Vercellotti, 2017; Yu & Lowie, 2020). Inspired by complex dynamic systems theory (CDST), recent research has increasingly focused on individual differences and the dynamics in CAF development (Li & Zhang, 2023; Vercellotti, 2017; Yu & Lowie, 2020). However, it is still a matter of controversy whether there are trade-off effects in CAF development, and how CAF relations change with time and differ across individual learners has not been fully explored. To address these gaps, this longitudinal study, guided by the CDST framework, attempted to explore individual differences and dynamic patterns in CAF relations by observing five Chinese undergraduates' English speech samples over 15 weeks.

2. Literature review

2.1. Definitions, hypotheses, and measures of CAF

It is first necessary to explain the basic constructs that were examined in this study. The construct of CAF was first proposed by Skehan (1996) to assess L2 ability in task-based experiments, and was later employed by some researchers to track second language (L2) development in longitudinal studies (Ferrari, 2012; Larsen-Freeman, 2006; Vercellotti, 2017; Yu & Lowie, 2020). CAF researchers generally agree that complexity refers to the degree of richness, diversity and sophistication in L2 syntactically and lexically, accuracy refers to native-like production in terms of grammar and lexical use, and fluency refers to target-like speed, pauses, repetitions and revisions in L2 speaking (Ellis, 2003, 2008; Ellis & Barkhuizen, 2005; Housen et al., 2012; Lennon, 1990; Michel, 2017; Skehan, 1996).

There are two key hypotheses regarding the relationship between complexity, accuracy and fluency. The limited attention capacity hypothesis, proposed by Skehan (1998), predicts that CAF components will compete with each other in tasks with higher cognitive demands. This hypothesis is built on the notion that one's attention capacity is limited; thus when a speaker focuses on one component, he or she will allocate less attention to other components (Skehan, 1998, 2014; Skehan et al., 2012), which is also called trade-off effects (Skehan, 1998). According to Skehan (1998), fluency competes with complexity and accuracy, and complexity also competes with accuracy. This hypothesis has been supported by some empirical evidence. Skehan and Foster (1997) found that fluency, accuracy and complexity seemed to compete with one another in tasks characterized by different cognitive demands, supporting not only the trade-off between meaning (fluency) and form (accuracy and complexity), but also the competition between the two aspects of form. The hypothesis has also been supported by some further empirical studies about the effects of task characteristics on CAF (Skehan, 2001, 2014; Skehan et al., 2012). These studies found that different task characteristics, including planning and repetition, enhanced learners' performance in one or two aspects of CAF, but never improved all three aspects, confirming the existence of trade-off effects in CAF relations. Other studies have also provided evidence for the competition between form and meaning. For example, Bygate (2001), when studying the effects of task repetition on oral performance, found that syntactic complexity competed with fluency, and Michel et al. (2007) observed competition between accuracy and fluency in a more complex task.

On the other hand, the cognition hypothesis, proposed by Robinson (2001, 2011), suggests that in certain tasks, complexity and accuracy may not necessarily trade off, and increases in task complexity along resource-directing dimensions could elicit more complex and accurate speech. The premise of this hypothesis is that there are multiple attention pools, allowing learners to allocate attention resources from different pools simultaneously (Robinson, 2001, 2011). In line with Skehan's (1998) hypothesis, Robinson (2001, 2011) also acknowledges that fluency competes with complexity and accuracy in oral tasks. The hypothesis has been supported by some empirical investigations. Yuan and Ellis (2003) found that students could achieve higher accuracy and complexity at the expense of fluency in conditions involving on-line planning. Likewise, Ahmadian and Tavakoli (2011) reported that complexity and accuracy competed with fluency in the condition of on-line planning, but accuracy and syntactic complexity increased together. Levkina and Gilabert (2012) observed an increase in complexity and accuracy at the cost of fluency in their studies on the effects of task complexity. A recent study by Awwad and Tavakoli (2022) also supported the simultaneous growth of syntactic complexity and accuracy in tasks involving a higher degree of intentional reasoning.

It can be concluded that both the limited attention capacity hypothesis and the cognition hypothesis support the trade-off effects between meaning (fluency) and form (accuracy and complexity). A meta-analysis by Jackson and Suethanapornkul (2013) found that task complexity had small positive effects on accuracy and small negative effects on fluency, partially supporting the trade-off effects. Another shared characteristic of these two hypotheses is that they discuss CAF relations under task effects rather than in second language development. In other words, these hypotheses were tested under experiment conditions, where one or two aspects of CAF increased with task complexity. Whether these hypotheses apply to the longitudinal development of L2 English speech is an issue that merits further exploration.

In CAF studies, one of the most challenging issues is measurement (Housen et al., 2012). Different studies may employ different CAF measures and no measure has been proved to be superior (Michel, 2017). However, certain measures are preferred over others for specific dimensions. For syntactic complexity, three commonly used measures are subordination, mean length of As-units and clauses per As-unit (Bulté & Housen, 2012). An As-unit is defined as an utterance that includes an independent clause or a sub-clause unit, along with any subordinate clauses that are linked to it (Foster et al., 2000). As-units are considered suitable units for measuring speech (Foster et al., 2000). When it comes to lexical complexity, the most popular measures are TTR (type-token ratio) and "D" (a measure of lexical diversity), with "D" being more effective in assessing speeches of various lengths (McKee et al., 2000). In terms of accuracy, widely used measures include error-free clauses and errors per 100 words (Ellis & Barkhuizen, 2005), while some local measures, such as error-free past tense forms, may also be used (Yu & Lowie, 2020). For fluency, researchers often employ measures such as speech rate, pause length and MLFR (mean length of fluent run) (Ellis & Barkhuizen, 2005). Different measures capture different aspects of CAF (Norris & Ortega, 2009), and CAF measures should be chosen in accordance with the purpose of the study and proficiency levels of subjects (Lambert & Kormos, 2014). Therefore, this study selected four general measures that are well-suited for undergraduates and effectively capture CAF relations.

2.2. CDST and CAF relations

From the perspective of CDST, second language development is viewed as a complex system characterized by dynamics, non-linearity, and openness to external influences (Larsen-Freeman, 1997, 2018, 2020; Lowie & Verspoor, 2022). Complex systems are composed of interactive subsystems and develop through the interaction of subsystems (every component can be considered as a subsystem) (Larsen-Freeman & Cameron, 2008; Lowie & Verspoor, 2022). The relationships between subsystems can be supportive, competitive, or even neutral (Bassano & van Geert, 2007; van Geert, 1994). However, it is important to note that these relationships are not stable or uniform (Larsen-Freeman, 1997, 2018, 2020; Lowie & Verspoor, 2022). From this perspective, variabilities, individual differences and dynamic interactions between components are seen as valuable information rather than mere noise (Larsen-Freeman, 1997, 2018, 2020; Lowie & Verspoor, 2022; Verspoor & de Bot, 2022; Verspoor et al., 2021).

Larsen-Freeman (2009) suggested that CDST could provide a broader framework for CAF studies. From this perspective, CAF is considered in terms of interrelated subsystems (Norris & Ortega, 2009) and two assumptions about CAF relations could be made. Firstly, CAF relations should differ from person to person because there is evidence that CAF trajectories are individualized (Larsen-Freeman, 2006; Yu & Lowie, 2020). Secondly, CAF relations should vary over time because CAF variables are developmental in nature (Norris & Ortega, 2009), and learners' performances exhibit significant variation over time (Larsen-Freeman, 2009; Yu & Lowie, 2020). As de Bot et al. (2007) predicted, trade-off effects might be found in CAF development, but these effects are likely to be unstable. A complex system is always dynamic, with varying degrees of variability, and a higher degree of variability is associated with systematic change (Larsen-Freeman & Cameron, 2008; Verspoor & de Bot, 2022; Verspoor et al., 2021). Therefore, it is necessary to identify the ways in which CAF relations change over time, that is, the dynamic patterns with different degrees of variability.

From the CDST perspective, longitudinal case studies tracking several individuals within a short time span (one semester) are meaningful for several reasons. Firstly, despite small sample sizes, longitudinal case studies are necessary due to the ergodicity problem in second language development. As Lowie and Verspoor (2019) stated, since "L2 learners do not form ergodic ensembles" and they "show clearly different learning trajectories over time," "longitudinal case studies are needed to understand the process of individual learners' development" (p. 184). Process-oriented research, which refers to research on how and why phenomena evolve or change with time (van Geert & Steenbeek, 2005), is the only way to discover the actual developmental process by focusing on numerous observations of individuals' language development (Lowie, 2017; van Geert & Steenbeek, 2005). Secondly, both short-term and long-term observations are significant in understanding the developmental process (van Geert & Steenbeek, 2005). Previous studies have shown that the development of CAF variables can be captured in short-term observations within the English as a foreign language (EFL) context by using general measures (Yu & Lowie, 2020). Thirdly, although measured variables may fluctuate from one time to another, CDST researchers deem fluctuation as information rather than measurement error and apply relevant methods to study the variability in development (van Geert & van Dijk, 2002).

2.3. Previous studies of CAF relations in L2 speech development

Research on CAF relations in L2 speech development has yielded valuable insights, primarily through longitudinal studies that track changes over time. The first longitudinal study in this field was conducted by Larsen-Freeman (2006), who adopted the perspective of CDST and viewed CAF in a dynamic manner. She collected the written and oral English production data (four times over six months) of five Chinese women who accompanied their husbands to study in the US. Using visual graphs, she observed that, overall, the group of five English learners made progress in CAF, but each individual followed a different path. Although the oral data were not analyzed as extensively as the written data, and the study did not specifically focus on CAF relations in oral English development, it brought the researchers' attention to the dynamic changes and individual differences in L2 CAF development.

Later, inspired by CDST, a small number of studies on L2 speech development emerged and confirmed the dynamic feature of CAF development (Ferrari, 2012; Li & Zhang, 2023; Polat & Kim, 2014; Vercellotti, 2017; Yu & Lowie, 2020). However, whether trade-off effects exist and how CAF relations change is still unclear. Some studies have found trade-off effects based on the overall trend of CAF development over a period of time. For instance, Ferrari (2012) conducted a three-year longitudinal study on four secondary-school Italian L2 learners and suggested a trade-off between complexity operationalized as mean length of As-units and accuracy measured by percentage of error-free As-units in multiple repeated tasks. Polat and Kim (2014) tracked accuracy, syntactic complexity, and lexical diversity in oral English interviews of a Turkish immigrant over one year. They found significant improvement in the untutored English learner's lexical diversity, measured by "D," but no development was observed in accuracy, measured by errors per 100 words. Li and Zhang (2023) conducted a longitudinal study tracking the English speech development of 45 Chinese undergraduates over a semester of an online course, collecting data six times. Latent growth curve modeling results revealed significant growth in accuracy, as measured by the percentage of error-free As-units, while no significant change was found in fluency, measured by the number of syllables per minute. These findings support the limited attention capacity hypothesis (Skehan, 1998), providing longitudinal evidence for trade-off effects between CAF variables in L2 development. When individuals focus on the improvement of one aspect over a period of time, other aspects receive less attention and may not grow or even experience a decline. These studies drew conclusions about CAF relations based on the overall trend during the observation period, but they did not utilize CDST methods to examine the dynamic relationship in the development process.

However, there are also studies that have found no trade-off effects. One of the most notable studies in this regard is the longitudinal study undertaken by

Vercellotti (2017). She utilized hierarchical linear and non-linear modeling techniques to examine English L2 oral data in topic-given monological tasks from 66 English as a second language (ESL) learners over a period of 3-10 months. Her findings indicated no trade-off effects between CAF variables measured by mean length of As-unit, "D," percentage of error-free clauses, and mean length of pause. Vercellotti's (2017) study is noteworthy for its large sample size and advanced analysis method, but still, the dynamic relationship during the process was not fully presented.

There are also other studies that have supported dynamic CAF relations. For example, Evans and Larsen-Freeman (2020) collected oral English data from an untutored adult French learner over one academic year, utilizing weekly topic-given monologic and dialogic tasks. They focused on syntax development and found that significant changes in the relationship between accuracy and fluency were indicative of systematic changes in syntax development. Another study, conducted by Yu and Lowie (2020), employed moving correlations (a method used to illustrate dynamic relationships between two variables) to analyze the relationship between complexity (measured by mean length of As-units and "D") and accuracy (measured by percentage of error-free As-units and error-free past tenses). They collected data from two Chinese college students over a period of twelve weeks, with weekly observations and monologue tasks on various topics. Their findings revealed that the relationship between complexity and accuracy exhibited a competitive nature in the early stages but became supportive in the later stages of development. These recent studies have partially validated dynamic CAF relations, but they did not cover all the three dimensions, and they only focused on one or two individuals, so individual differences and dynamic patterns were not fully explored.

As can be seen from this overview, there is a scarcity of longitudinal studies examining CAF relations in L2 speech development, and the presence of trade-off effects between CAF variables remains a subject of debate. Furthermore, while some previous studies have concentrated on individual differences and dynamic changes in CAF development (Ferrari, 2012; Larsen-Freeman, 2006; Yu & Lowie, 2020), they have not adequately explored individual differences and dynamic patterns in CAF relations, which should be highlighted from the perspective of CDST.

3. The study

3.1. Aims and research questions

Informed by CDST, the current study aims to validate the assumption of individualized and dynamic CAF relations in L2 speech development (de Bot et al., 2007;

Larsen-Freeman, 2006; Yu & Lowie, 2020), and identify dynamic patterns in CAF relations. This longitudinal study was conducted to answer the following research questions:

1. How do CAF relations differ across different individuals? (RQ1)
2. What are the dynamic patterns in CAF relations? (RQ2)

3.2. Participants

In process-oriented research, participants are not necessarily statistically representative, but they should belong to the same class in terms of age and learning context (van Geert & Steenbeek, 2005). Following this principle, the convenience sampling method was employed in selecting five participants (aged 20-21) who were sophomores at a Chinese university. S1, S4 and S5 were male, and S2 and S3 were female. All participants were studying English as their major and were enrolled in the same English courses, including reading, writing, listening and speaking. The speaking course, taught by a native speaker, took place twice a week for 180 minutes, with the aim of improving the students' overall speaking ability. No specific training focusing on complexity, accuracy, or fluency was provided. Outside the classroom, the participants rarely engaged in English conversation among themselves. Throughout the observation period, these participants resided in China and had no prior experience of living in or visiting any English-speaking country.

3.3. Research procedures

3.3.1. Data collection and transcription

The study involved 15 distinct oral monologic tasks. Each of the five participants visited the lab individually to perform one task each week for 15 consecutive weeks. Participants were given one minute of preparation time without note-taking and then spoke for approximately three minutes, with their speeches recorded using a voice recorder. Each of these 15 tasks required participants to provide narratives about personal experiences, with consistent cognitive demands across all topics. Examples of the task prompts included "Describe a lesson you have learned which has enriched your life experience" or "Talk about an animal that has brought you an unforgettable experience." All of the topics were selected from the TEM-4 spoken test, which is an examination for English majors in China taken by sophomores. The participants had not practiced these specific topics beforehand. To ensure comparable levels of topic difficulty, the 15 topics were rated by 72 second-grade English majors (excluding the five participants). The 15 topics had similar difficulty levels, with scores ranging between 2.53 and 2.88 on a scale of 1 to 5, with 1 indicating very easy and 5 indicating very difficult.

All the oral recordings were then transcribed by the first author according to CHAT transcription format (MacWhinney, 2000), and CLAN, a specialized tool for talk analysis, was applied to transcribe and analyze all the oral data. In the transcripts, utterances were segmented into As-units according to established principles (Foster at al., 2000) for the convenience of CAF analysis. After transcription, two speech samples were randomly chosen from each participant's 15 samples (more than 10% of all the samples) and checked by the second author. The inter-transcriber agreement reached 97.3%, indicating a high level of reliability in the transcription process.

3.3.2. CAF measures

The current study adopted four general measures to track the development of complexity, accuracy and fluency. Complexity was measured both syntactically and lexically (Bulté & Housen, 2012). The syntactic complexity measure was "mean length of As-unit" as it has been shown to be sensitive to short-term development in the English speech of Chinese undergraduates (Yu & Lowie, 2020). This measure counted the average number of words per As-unit, excluding repetitions and self-revisions. Given that the transcripts were already divided into As-units, the mean length of As-unit could be easily computed using the CLAN software. The lexical complexity measure was "D," which has been claimed to be more valid than TTR when measuring speeches of various lengths (McKee et al., 2000). Therefore, "D" was deemed suitable for analyzing the current dataset and could be conveniently computed using CLAN.

Accuracy was measured by "errors per 100 words," excluding repetitions and self-revisions (Ellis & Barkhuizen 2005). This measure was chosen for two main reasons. Firstly, it was not affected by the division of clauses (Ellis & Barkhuizen, 2005), unlike measures such as the "percentage of error-free clauses" and the "weighted clause ratio" (Foster & Wigglesworth, 2016). Secondly, it was more suitable for our participants compared to the "percentage of error-free As-units" since some learners tended to make multiple errors within a single As-unit. Errors related to both grammar and vocabulary were calculated, and they were determined by the two authors based on English grammatical rules and the Oxford English Dictionary. To ensure consistency, an initial sample of ten recordings was randomly selected and independently coded by both raters, resulting in an inter-rater agreement of 94%. Subsequently, the remaining samples were divided into two groups and each rater coded their respective group. The two raters then cross-checked each other's coding, consulted a native speaker in case of disagreements, and reached a consensus on every coded error.

Fluency was measured by MLFR (mean length of fluent run), which has been demonstrated to be an effective indicator of fluency (Kormos, 2006; Tonkyn, 2012; Towell et al., 1996). Consistent with Tonkyn (2012), we considered sequences of syllables between pauses lasting over 0.3 seconds as fluent runs, because pauses shorter than 0.3 seconds may not represent real hesitations for our participants. To facilitate the calculation of MLFR, the speech analysis software Praat was utilized.

3.3.3. Data analysis

We followed the methods outlined by Verspoor and van Dijk (2011) to analyze CAF relations. Firstly, we normalized the CAF data for each individual using the formula (x-min)/(max-min), which transformed all the data onto a common scale ranging from 0 to 1 (Verspoor & van Dijk, 2011). Next, we employed a smoothing technique called LOWESS in the software STATA to visualize the interactions between variables. When using LOWESS, we experimented with different bandwidths and ultimately selected a bandwidth of 0.6, as it yielded relatively clear relations while preserving interesting variability. Following Verspoor and van Dijk (2011), we re-normalized the smoothed data to enhance the clarity of the interactions.

Correlation analysis was then conducted on each individual's normalized CAF data to validate noticeable relations observed in the visual graphs. Finally, "moving correlations" (Verspoor & van Dijk, 2011), a technique to present dynamic relations between two variables, was applied to help identify dynamic patterns of CAF relations. In our study, a window of five observations was used, meaning that we calculated correlation coefficients between two normalized variables from Week 1 to Week 5, Week 2 to Week 6, and so on, up to Week 11 to Week 15. These coefficients were then utilized to construct moving correlation graphs. For a comprehensive understanding of the procedures, please refer to Verspoor and van Dijk (2011). In identifying patterns, we followed the effect size guidelines suggested by Plonsky and Oswald (2014) to assess the relationship: neutral ($-.25 < r < .25$), competitive ($r \leq -.25$), and supportive ($r \geq .25$).

4. Findings

4.1. Individual differences in CAF relations

The descriptive statistics of five individuals' CAF data are provided first to offer an overview of their individual CAF levels. Subsequently, the smoothed and normalized trajectories of each participant's CAF variables, along with the corresponding correlation analysis results, are presented.

4.1.1. Descriptive statistics of individuals' CAF data

Table 1 displays the initial levels and means of each participant's CAF variables across the 15-week period. The data show that among the five learners, S1 exhibited the highest initial and average levels of lexical complexity. S2 had the lowest initial level of lexical complexity and fluency. S3 was featured by the highest initial and average levels of syntactic complexity. S4 demonstrated the highest initial level of fluency but the lowest initial and average levels of accuracy. Finally, S5 was characterized by the highest initial level of accuracy but the lowest initial level of syntactic complexity.

Table 1 Descriptive statistics of individuals' CAF data over 15 weeks

Participants	C1		C2		A		F	
	Initial level	*M*(*SD*)	Initial level	*M*(*SD*)	Initial level	*M*(*SD*)	Initial level	*M*(*SD*)
S1	9.05	9.26 (1.45)	68.62	49.22 (10.68)	11.63	4.98 (2.58)	4.13	4.01 (.61)
S2	11.12	10.21 (1.51)	24.77	37.60 (7.41)	6.88	8.52 (2.47)	2.74	2.42 (.27)
S3	13.88	10.84 (1.58)	36.48	43.92 (8.41)	4.66	3.99 (1.48)	3.75	3.78 (.50)
S4	9.05	8.91 (1.09)	39.03	37.10 (6.26)	15.08	13.40 (2.97)	6.38	4.09 (1.17)
S5	8.22	9.12 (1.20)	46.68	42.68 (5.60)	4.05	5.64 (2.26)	5.28	4.80 (.54)

Note. C1 = mean length of As-unit; C2 = "D;" A = errors per 100 words; F = MLFR.

4.1.2. CAF relations in S1's speech development

Figure 1 visualizes the smoothed and normalized trajectories of the CAF variables over the 15-week period. It is evident that the learner improved his accuracy throughout the entire duration, while his lexical complexity and fluency showed a decline. Notably, the trajectories of the lexical complexity measure "D" and the accuracy measure "errors per 100 words" exhibited a high degree of similarity, indicating a relationship between lexical complexity and errors. This trade-off between accuracy and lexical complexity was further supported by correlation analysis (see Table 2). Furthermore, Figure 1 illustrates that the development of the syntactic complexity measure "mean length of As-unit" and the fluency measure MLFR generally followed similar trends, with the exception of three instances in Weeks 7-8, 10-11, and 14-15. This overall correlation between syntactic complexity and fluency was confirmed by correlation analysis (see Table 2). Moreover, the development of "errors per 100 words" and MLFR primarily exhibited opposite directions from Week 4 onwards. Correlation analysis showed an overall negative correlation between A and F (see Table 2), indicating a supportive relation between accuracy and fluency.

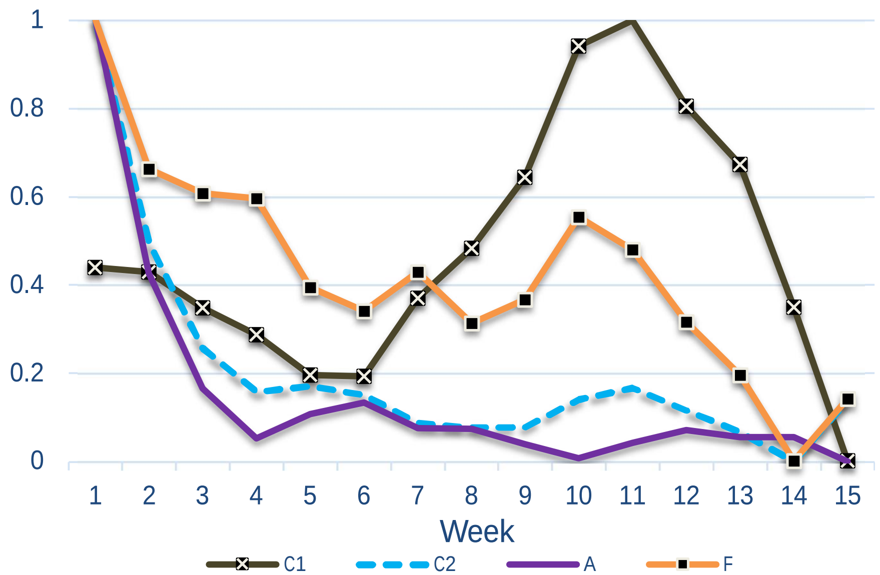

Figure 1 Smoothed and normalized trajectories of S1's CAF variables (C1 = mean length of As-unit; C2 = "D;" A = errors per 100 words; F = MLFR)

Table 2 Correlation matrix of S1's CAF variables

	C1	C2	A
C2	.276		
A	-.343	.566*	
F	.536*	.035	-.534*

Note. C1 = mean length of As-unit; C2 = "D;" A = errors per 100 words; F = MLFR; * $p < .05$

4.1.3. CAF relations in S2's speech development

Figure 2 shows that S2 made progress in lexical complexity, while her accuracy and fluency became lower. The trajectories of "D" and "errors per 100 words" exhibited similarity from Week 6 to 14, and this association was statistically significant according to the correlation analysis presented in Table 3. This indicates an overall competitive relationship between lexical complexity and accuracy. Initially, the relationship between these variables was not clear, but as time progressed, S2's improvement in lexical complexity came at the cost of accuracy.

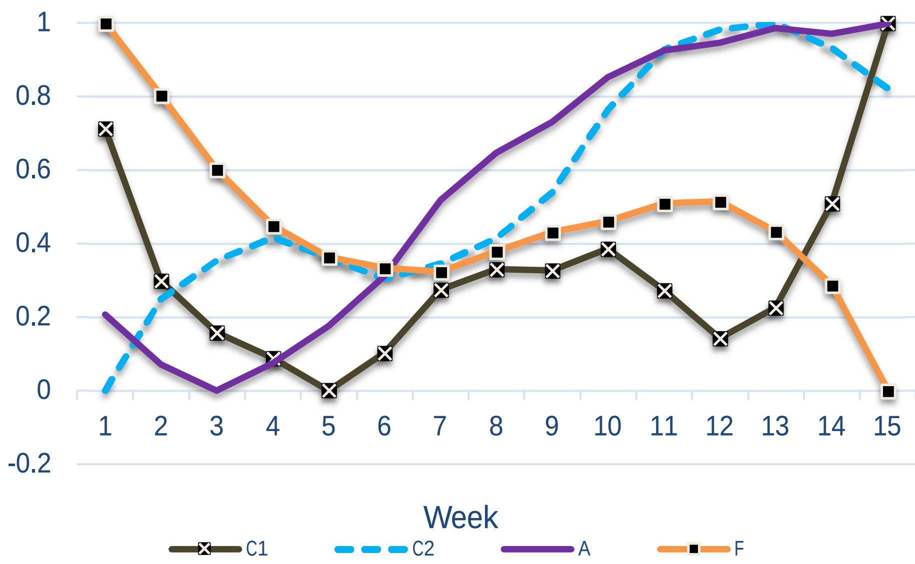

Figure 2 Smoothed and normalized trajectories of S2's CAF variables (C1 = mean length of As-unit; C2 = "D;" A = errors per 100 words; F = MLFR)

Table 3 Correlation matrix of S2's CAF variables

	C1	C2	A
C2	.179		
A	.184	.684**	
F	.235	.170	-.400

Note. C1 = mean length of As-unit; C2 = "D;" A = errors per 100 words; F = MLFR; ** *p* < .01

4.1.4. CAF relations in S3's speech development

Figure 3 shows that S3 improved in lexical complexity and accuracy, while her syntactic complexity and fluency decreased. The trajectories of "errors per 100 words" and MLFR were similar, suggesting a competitive relationship between accuracy and fluency. Meanwhile, the development of "mean length of As-unit" and "D" showed opposite directions for the most part, indicating a competition between syntactic complexity and lexical complexity. These overall competitions between the variables were further confirmed by the correlation analysis presented in Table 4.

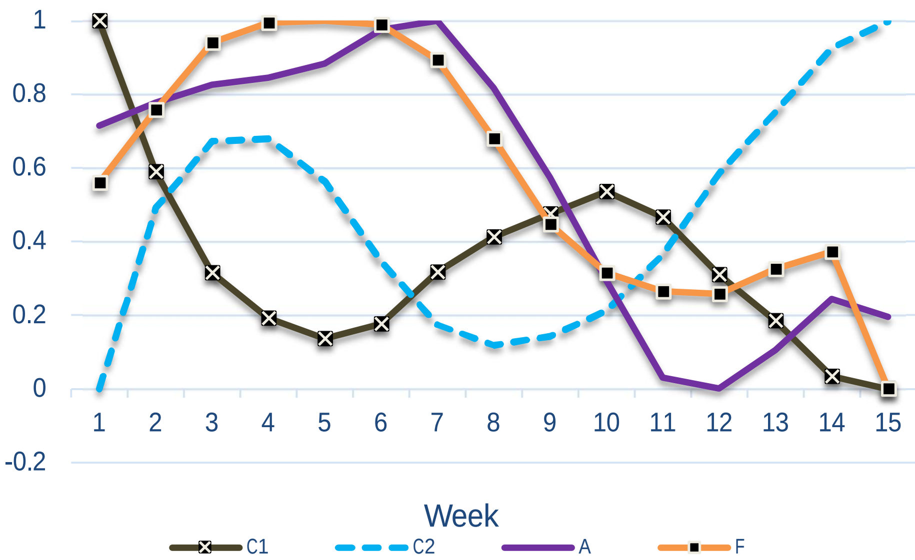

Figure 3 Smoothed and normalized trajectories of S3's CAF variables (C1 = mean length of As-unit; C2 = "D;" A = errors per 100 words; F = MLFR)

Table 4 Correlation matrix of S3's CAF variables

	C1	C2	A
C2	-.550*		
A	-.051	.346	
F	-.321	.256	.667**

Note. C1 = mean length of As-unit; C2 = "D;" A = errors per 100 words; F = MLFR; * $p < .05$. ** $p < .01$

4.1.5. CAF relations in S4's speech development

The CAF trajectories depicted in Figure 4 reveal a noticeable increase in S4's syntactic complexity and a clear decrease in fluency. Although the fluctuation of the "mean length of As-unit" was more pronounced than that of the MLFR, it is evident that these two variables developed in opposite directions from Week 3 onwards, indicating an overall competition between syntactic complexity and fluency. The correlations between the CAF variables (see Table 5) also supported this competitive relationship.

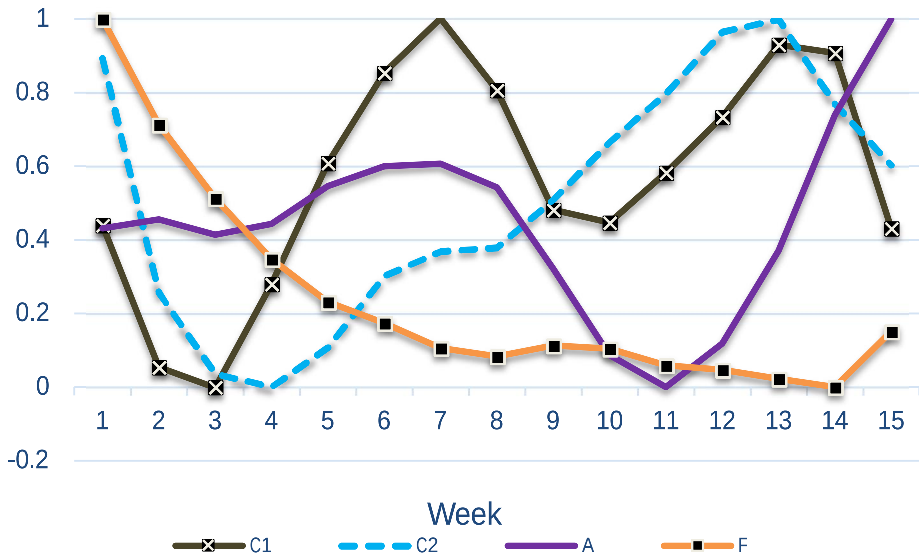

Figure 4 Smoothed and normalized trajectories of S4's CAF variables (C1 = mean length of As-unit; C2 = "D;" A = errors per 100 words; F = MLFR)

Table 5 Correlation matrix of S4's CAF variables

	C1	C2	A
C2	.265		
A	.138	-.261	
F	-.602*	.130	-.056

Note. C1 = mean length of As-unit; C2 = "D;" A = errors per 100 words; F = MLFR; * *p* < .05.

4.1.6. CAF relations in S5's speech development

Figure 5 illustrates that S5 improved in syntactic complexity and experienced a decline in fluency and accuracy. However, the correlation analysis (see Table 6) did not reveal any significant relationships between the CAF variables.

Table 6 Correlation matrix of S5's CAF variables

	C1	C2	A
C2	-.126		
A	-.482	-.225	
F	-.312	.480	.126

Note. C1 = mean length of As-unit; C2 = "D;" A = errors per 100 words; F = MLFR.

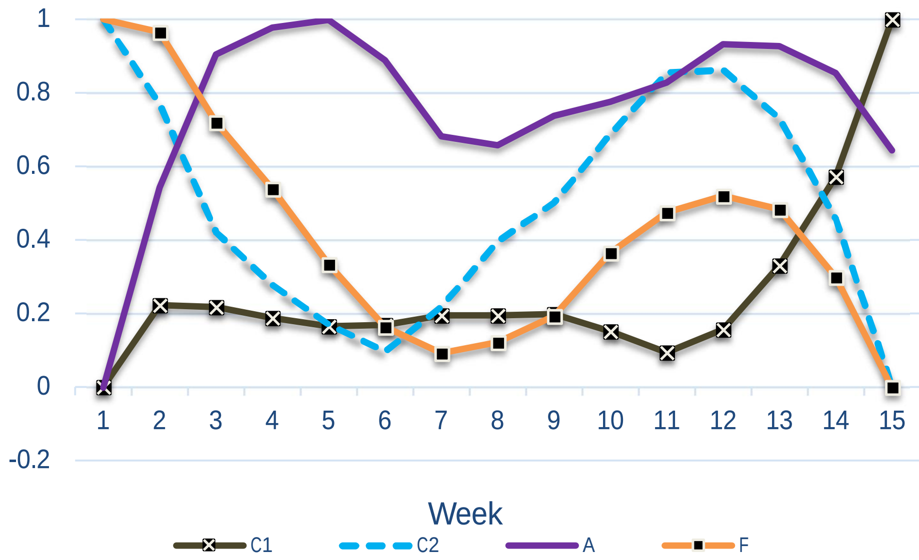

Figure 5 Smoothed and normalized trajectories of S5's CAF variables (C1 = mean length of As-unit; C2 = "D;" A = errors per 100 words; F = MLFR)

4.2. Dynamic patterns in CAF relations

By observing moving correlations of individuals' CAF correlation coefficients (with 5 observations per window), we identified eight distinct patterns of change in CAF relations (see Table 7). They are Pattern A: "competitive to supportive," Pattern B: "supportive to competitive," Pattern C: "competitive to neutral," Pattern D: "supportive to neutral," Pattern E: "neutral to competitive," Pattern F: "neutral to supportive," Pattern G: "stable with occasional changes," and Pattern H: "irregular changes." Each of these patterns will be further elucidated and exemplified using specific samples.

Pattern A ("competitive to supportive") indicates a shift in the relationship from competitive to supportive during the 15-week observation period. An example of Pattern A is demonstrated in Figure 6. In this example, the correlation between S1's lexical complexity and fluency changed from negative to positive. Although the overall correlation between these two aspects, as indicated in Table 2, was not significant, the moving correlations offer more detailed insights into the dynamic nature of the relationship.

Pattern B ("supportive to competitive") signifies a transition in the relationship from supportive to competitive within the observation period. This pattern is exemplified in Figure 7. Since the accuracy measure "errors per 100 words" was a negative indicator, correlation coefficients were reversed to present the relationship between accuracy and lexical complexity. Specifically, positive correlation coefficients were reversed to negative, and negative correlation coefficients were reversed to positive. This reversal allows for a clear understanding of the relationship between accuracy and lexical complexity. As depicted in Figure 7, S5's lexical complexity initially exhibited a positive correlation with accuracy during the early stage when both aspects decreased (as observed in Figure 5). However, in the later stage, the correlation shifted to negative as lexical complexity increased. This shift coincided with the increase in lexical complexity. This pattern tended to occur when one of the two factors experienced a continuous increase throughout the process.

Table 7 Dynamic patterns in CAF relations

Dynamic patterns	Syntactic complexity and lexical complexity	Syntactic complexity and accuracy	Syntactic complexity and fluency	Lexical complexity and accuracy	Lexical complexity and fluency	Accuracy and fluency
A	S1		S5		S1	S4
B				S5		
C		S2	S3	S1		
D		S1, S5				
E	S3		S2	S2		
F		S3				
G			S1, S4		S5	S1, S2, S3
H	S2, S4, S5	S4		S3, S4	S2, S3, S4	S5

Note. A = "competitive to supportive;" B = "supportive to competitive;" C = "competitive to neutral;" D = "supportive to neutral;" E = "neutral to competitive;" F = "neutral to supportive;" G = "stable with occasional changes;" H = "irregular changes".

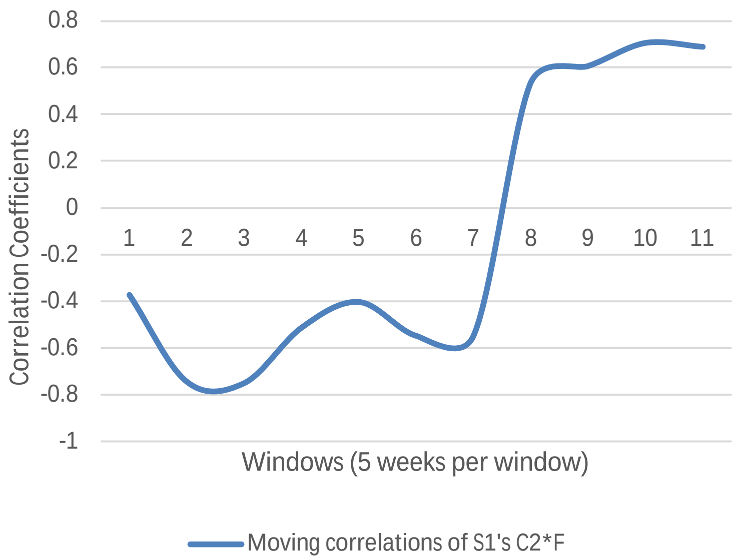

Figure 6 Sample of Pattern A: "competitive to supportive" (C2 = "D;" F = MLFR)

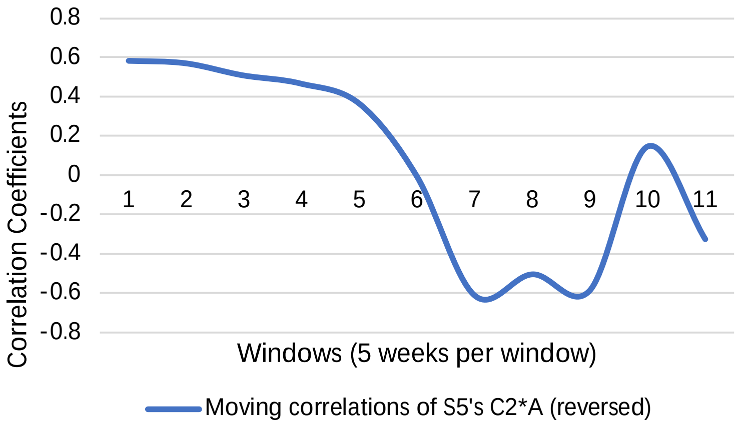

Figure 7 Sample of Pattern B: "supportive to competitive" (C2 = "D;" A = errors per 100 words)

Pattern C ("competitive to neutral") shows that the competition became less evident with time, and the relationship became neutral during the 15 weeks. An example is presented in Figure 8. The competition between S1's lexical complexity and accuracy became less evident with time. As depicted in Figure 1, accuracy increased a lot in the early stage, leading to a decrease in lexical complexity due to the trade-off between the two factors. However, in the later stage, when both aspects remained relatively stable, the competition between them disappeared, resulting in a neutral relationship.

Pattern D ("supportive to neutral") indicates a gradual weakening of a positive relationship over time, leading to a neutral relationship during the observation period. To illustrate this pattern, we can examine S1's syntactic complexity and accuracy as an example (see Figure 9). In the early stage, there was a clear positive correlation between these two constructs, suggesting a supportive relationship. However, as time progressed, this positive correlation became weaker. By the end of the observation period, the relationship had become neutral. The trend observed in the data implies that accuracy and syntactic complexity may potentially compete with each other beyond the observation period.

Pattern E ("neutral to competitive") describes a transition from a neutral relationship to a competitive one. It is exemplified by S2's lexical complexity and accuracy (see Figure 10). Initially, a neutral relationship existed between these two aspects. However, over time, the neutral relationship disappeared, and the two aspects became increasingly competitive. The trade-off effects between these two aspects remained relatively stable and were likely to persist beyond the observation period. This competition was caused by the continuous increase in lexical complexity, as indicated in Figure 2.

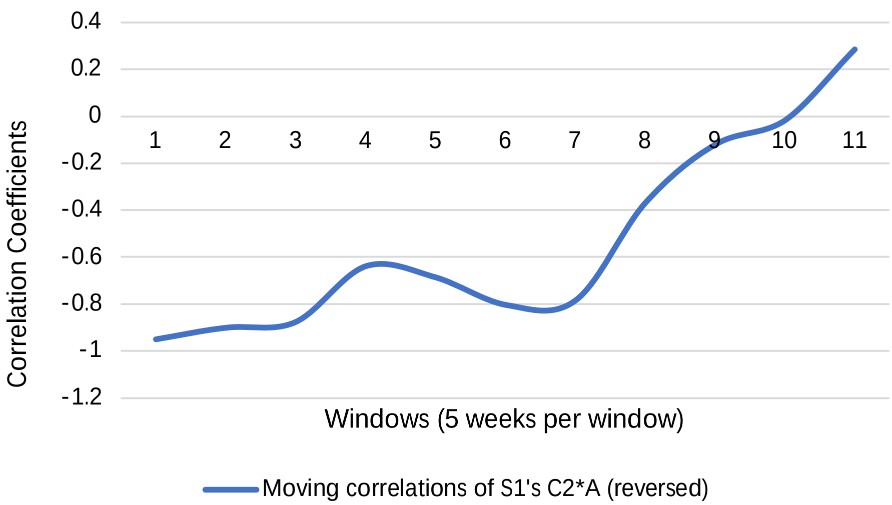

Figure 8 Sample of Pattern C: "competitive to neutral" (C2 = "D;" A = errors per 100 words.

Pattern F ("neutral to supportive") refers to situations where a neutral relationship turned into a supportive one. An example is illustrated in Figure 11. The neutral relationship between S3's accuracy and syntactic complexity became a supportive relationship later. As previously shown in Figure 3, during the 15 weeks, S3's accuracy improved, while syntactic complexity decreased. However, these two aspects were not in direct competition with each other. Instead, accuracy competed with fluency, and syntactic complexity competed with lexical complexity.

Pattern G ("stable with occasional changes") indicates that the relationship between variables did not undergo significant overall changes but experienced occasional fluctuations in correlation coefficients within one or two windows during the process. This pattern is demonstrated by S1's syntactic complexity and fluency in Figure 12. Except for an occasional change, the correlation level remained relatively high during the observation period.

Pattern H ("irregular changes") signifies that the relationship between variables underwent considerable and unpredictable fluctuations with no clear trend during the observation period. This pattern emerged as the most frequent one in our study. An example illustrating this pattern is demonstrated by the relationship between S5's syntactic complexity and lexical complexity, as depicted in Figure 13. As illustrated, the relationship between these two aspects was complex and unpredictable, characterized by irregular changes and lacking a discernible trend.

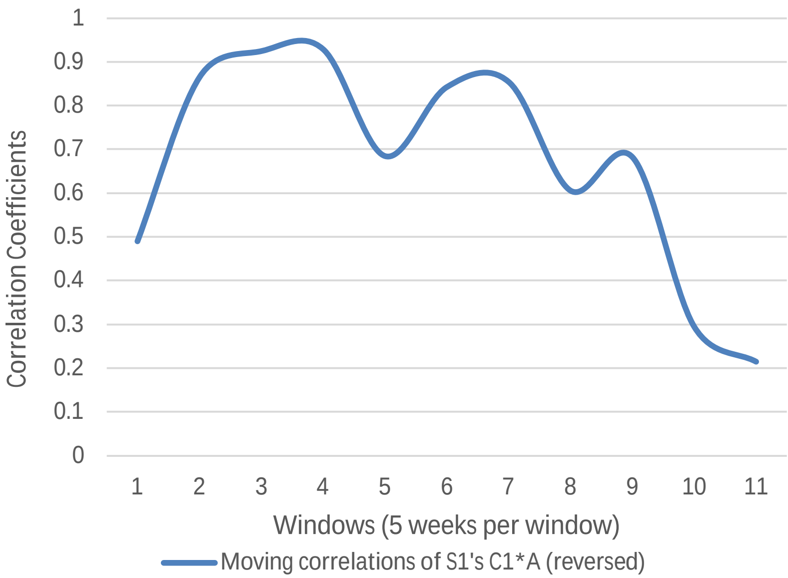

Figure 9 Sample of Pattern D: “supportive to neutral” (C1 = mean length of As-unit; A = errors per 100 words)

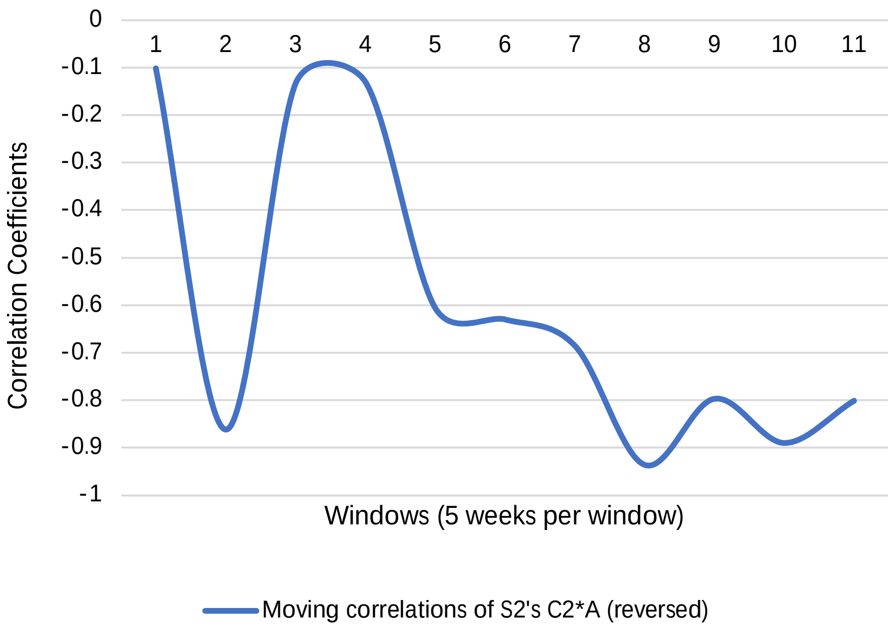

Figure 10 Sample of Pattern E: “neutral to competitive” (C2 = “D;” A = errors per 100 words)

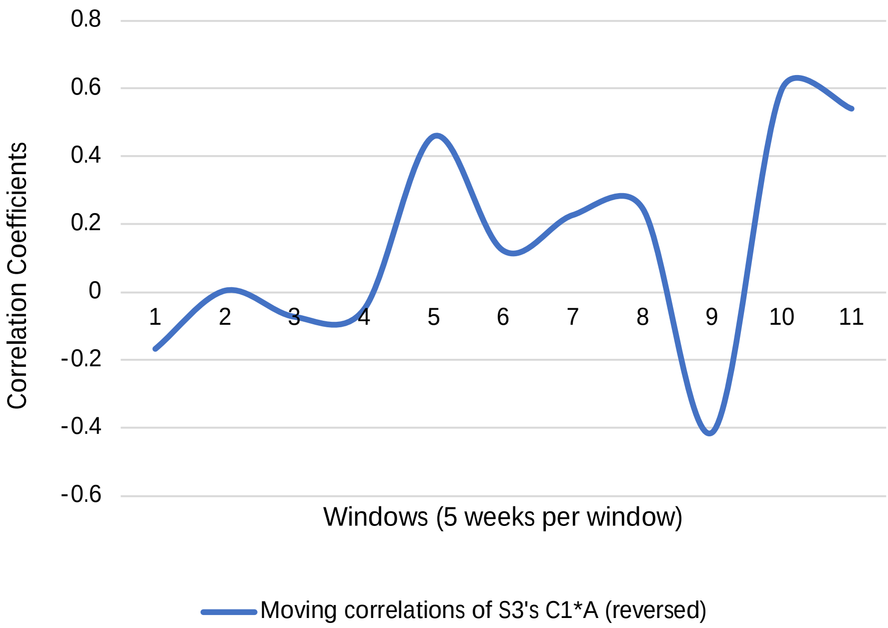

Figure 11 Sample of Pattern F: “neutral to supportive” (C1 = mean length of As-unit; A = errors per 100 words)

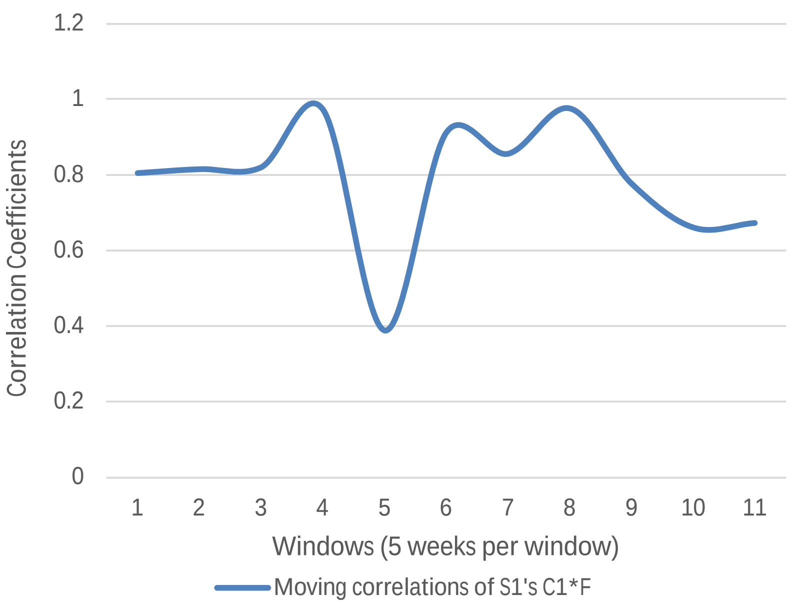

Figure 12 Sample of Pattern G: “stable with occasional changes” (C1 = mean length of As-unit; F = MLFR)

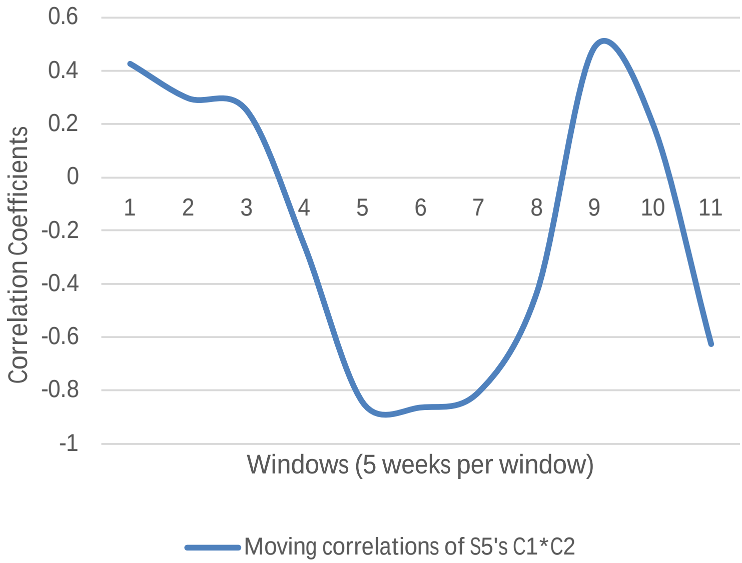

Figure 13 Sample of Pattern H: "irregular changes" (C1 = mean length of As-unit; C2 = "D")

5. Discussion

The findings of this study affirm the hypothesis of individualized and dynamic CAF relations in L2 speech development (de Bot et al., 2007; Larsen-Freeman, 2006; Yu & Lowie, 2020). In the following two sections, we will discuss the observed individual differences and dynamic patterns in CAF relations, respectively.

5.1. Individual differences in CAF relations (RQ1)

LOWESS graphs and correlation analyses showed that there existed individual differences in CAF relations. Different trade-off effects were found in the CAF development of four individuals. One learner improved in lexical complexity at the expense of accuracy, another learner improved in accuracy at the cost of lexical complexity, one student made progress in syntactic complexity but sacrificed fluency, and one improved in accuracy and lexical complexity while compromising fluency and syntactic complexity, respectively.

This longitudinal study supports the CDST perspective that individual differences in language development do exist and should not be overlooked (Larsen-Freeman, 2006; Lowie & Verspoor, 2019). The current study expanded

Larsen-Freeman's (2006) study of CAF development in L2 writing to CAF relations in L2 speaking. Our data indicate that in oral English development, individuals also followed different paths, and CAF relations differed across learners. It was observed that different individuals may prioritize different dimensions of CAF over a given period, and trade-off effects were evident among some of the undergraduate participants.

Some trade-off effects found in individual learners' CAF development align with previous studies. Firstly, the trade-off between lexical complexity and accuracy (for S2) was consistent with the findings of Polat and Kim's (2014) longitudinal study on untutored immigrants. In their study, lexical complexity showed significant improvement, while accuracy did not progress significantly. Our study extends these findings by demonstrating that the trade-off between lexical complexity and accuracy can also occur in the EFL context. Secondly, the trade-off between accuracy and fluency (for S3) has been previously reported in some studies investigating task effects (Ahmadian & Tavakoli, 2011; Levkina & Gilabert, 2012; Michel et al., 2007; Skehan & Foster, 1997; Yuan & Ellis, 2003). The current study provided longitudinal evidence that certain EFL learners may improve their oral accuracy at the cost of oral fluency. Lastly, the competition between syntactic complexity and fluency (for S4) has been documented in many studies of task effects (Ahmadian & Tavakoli, 2011; Bygate, 2001; Levkina & Gilabert, 2012; Skehan & Foster, 1997; Tonkyn, 2012). Our study adds to the existing literature by presenting new longitudinal evidence of trade-off effects between syntactic complexity and fluency.

However, our findings diverged from those reported by Vercellotti (2017), who found that students did not prioritize the development of one CAF construct at the expense of another. This disparity may be attributed to two factors. Firstly, the difference in learning context may partially account for the divergent findings. The participants in the study of Vercellotti (2017) were enrolled in a language training program in the target-language country, while our study observed L2 learners in their native country. The language context plays a vital role in one's second language development (Taguchi, 2013). As mentioned earlier, our participants took the speaking course twice a week and had limited opportunities to communicate in English outside the classroom. In contrast, the participants in Vercellotti's (2017) study attended speaking courses more frequently and used English to communicate beyond the classroom. This difference in learning context may have led to substantial discrepancies in English input and output, ultimately influencing the rate of development. Another reason for the divergence could be the difference in the duration of observation. Our study tracked L2 speech development over 15 weeks, while Vercellotti (2017) collected oral data over a span of 3-10 months. It was also confirmed in our study that trade-off in the early stage might turn into a neutral or supportive relationship in the later stages. Consequently, data collected over a longer observation period could potentially yield different conclusions.

Although the differences between the EFL learners in this process-oriented study may not be statistically representative, the participants share certain common characteristics, as van Geert and Steenbeek (2005) suggested. For instance, they studied in the same context in terms of culture, language and curriculum. In this study, we found that, despite the variability and complexity in oral English development, it is also possible to identify some common features in CAF development and relations. One notable finding in our study was that all learners experienced a decline in fluency to varying degrees. This suggests that fluency received less attention in the oral English development of these Chinese undergraduate students who did not receive specialized training in complexity, accuracy, or fluency. This shared trend highlights the need to address fluency issues in language learning and teaching contexts, particularly among learners who may not prioritize it naturally.

Our study also supports the CDST view that initial conditions have a considerable effect on the development of complex systems (Larsen-Freeman & Cameron, 2008). Individual differences in CAF relations can be attributed to various factors, including the initial level of CAF. In our study, we observed that the initial level of CAF played a crucial role in shaping the development of individual learners. For instance, S2's initial level of lexical complexity was the lowest, and she improved in this aspect at the expense of accuracy. On the other hand, S4's initial level of fluency was the highest among all participants, but his syntactic complexity was very low at the beginning. He experienced great improvement in syntactic complexity during the observation period, at the cost of fluency. Furthermore, S5, who already had a notably high level of accuracy and relatively high fluency at the outset, did not demonstrate significant improvement during the 15-week period. These examples illustrate how the initial level of CAF can contribute to individual differences in CAF development and relations.

5.2. Dynamic patterns in CAF relations (RQ2)

The eight change patterns of CAF relations support the hypothesis that CAF relations are subject to variation over time. The current study partially aligns with previous research in this area. For example, we identified a pattern called "competitive to supportive," which indicates that the competition between CAF components in the early stages turned into a supportive relationship in later stages. This finding is in line with the conclusions drawn by Yu and Lowie (2020) regarding complexity and accuracy. Furthermore, our study goes beyond previous research by identifying additional patterns through the examination of multiple learners' development across all three dimensions of CAF.

Our findings about the dynamic patterns in CAF relations contribute to and advance the CDST in three key aspects. Firstly, as Larsen-Freeman (2009) emphasized, when examining CAF dimensions, we should not "miss their interaction and the fact that the way that they interact changes with time as well" (p. 582). In line with this viewpoint, the current study not only provided evidence for CAF interaction but also demonstrated the importance of understanding CAF relations "that change throughout the course of development," rather than "static or one-time relations between variables" (Larsen-Freeman, 2009, p. 583). Longitudinal and individual case studies do not aim to arrive at definitive conclusions about CAF relations that apply universally. By contrast, we prioritized the dynamic patterns exhibited by individuals, and this process-oriented study enabled us to obtain a deeper understanding of CAF relations in terms of their dynamics, as well as identify common features in the developmental process.

Secondly, the eight change patterns of CAF relations confirmed that variability is common in development and a higher degree of variability is related to systematic change (Verspoor & de Bot, 2022; Verspoor et al., 2021). Among all the patterns, patterns A and B presented the highest variability, and indicated a dramatic shift from competitive to supportive or vice versa. Patterns C and D demonstrated medium variability, and they "reflected potential for more dramatic change." Patterns E and F showed low-medium variability, and they suggested potential for "long-term stability" (Cameron & Larsen-Freeman, 2007, p. 229). Pattern G exhibited the lowest variability and kept relatively stable. Although pattern H did not reveal meaningful trends during the observation, it may show regularity if the observation period was extended.

More importantly, this study confirmed the dynamic nature of trade-off effects. Our findings support the prediction that there might be trade-off effects between CAF constructs in oral English development, but the relationship is likely to vary over time (de Bot et al., 2007). Our study shows that trade-off effects indeed took place, and the competition was not limited to fluency versus the other two constructs, but also extended to complexity versus accuracy. This is consistent with the limited attention capacity hypothesis (Skehan, 1998). Additionally, our study reveals that trade-off effects change according to the developmental stage of CAF variables. Based on our analysis, competition typically arose when one of the two variables experienced continuous growth and might disappear or transform into a supportive relationship when no variable exhibited continuous growth. This implies that cross-sectional studies may fail to capture the dynamic changes of CAF relations and ignore the complexity of CAF relations. The relationship between CAF constructs should not be oversimplified as merely “competing” or “supporting” without acknowledging their dynamic nature.

To sum up, this study highlighted the importance of variability and individual differences and employed longitudinal data to validate the perspective of CDST that CAF relations in L2 English speech development are dynamic and individualized. The findings confirmed the existence of trade-off effects, but emphasized the need for a dynamic view when interpreting them.

6. Conclusions

According to this longitudinal study, CAF relations vary from person to person and are dynamic from time to time. Individual learners receiving the same instructions showed individual differences in CAF relations. Different trade-off effects were found in the CAF development of four individuals: One learner improved in lexical complexity at the expense of accuracy, another learner improved in accuracy at the cost of lexical complexity, one student made progress in syntactic complexity but sacrificed fluency, and one improved in accuracy and lexical complexity while compromising fluency and syntactic complexity, respectively. Furthermore, CAF relations changed with time dynamically and eight change patterns were identified. These patterns exhibited varying degrees of variability, ranging from highest to medium to low-medium to lowest. Six patterns involved state changes, transitioning from a competitive to supportive relationship, for example.

The results of this study have significant implications for L2 English teaching and learning. One implication is that learners who have not received specific training focusing on CAF may inadvertently prioritize other aspects of language development over fluency. In light of this, teachers can play a crucial role in guiding students to allocate conscious attention to fluency enhancement. This can be achieved by encouraging students to avoid unnecessary repetitions and reformulations in their speech. By consciously attending to fluency while simultaneously working on accuracy or complexity, students may experience overall improvements in their L2 speech performance. Another pedagogical implication relates to the recognition that students within the same class may be at different stages of CAF development. Consequently, it is important for both teachers and students to be aware of individual differences and characteristics when setting language learning goals. Recognizing and addressing these individual features can help tailor instruction to meet the specific needs and priorities of each student, thereby fostering more effective and personalized language learning experiences.

However, it is important to acknowledge the limitations of this study. Although we investigated the variations in CAF relations across time and learners, we only attributed the differences to the individuals' initial level of CAF. It is evident that there are additional factors influencing CAF development and relations both internally (such as motivation and engagement) and externally (such as the learning context). These factors should be explored in a dynamic manner to gain a more comprehensive understanding of CAF. This study was also constrained by a small sample size, which restricted our ability to employ more advanced modeling techniques for analyzing individual differences and dynamic patterns. Additionally, the limited number of CAF measures utilized in this study may not have captured the complete spectrum of CAF relations. To address these limitations, future research should incorporate a broader range of factors, involve a larger number of participants, and employ multidimensional measures to provide a more nuanced understanding of CAF development and relations.

Acknowledgements:

Project of Humanities and Social Sciences funded by the Education Department of Jilin Province, China (Project No. JJKH20241400SK).

We would like to express our sincere gratitude to the anonymous reviewers and the Editor in Chief for their valuable feedback and insightful suggestions. We are also very grateful to Professor Irene Hanson Frieze for her invaluable assistance, which enhanced the clarity and coherence of our work.

References

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research, 15*(1), 35-59. https://doi.org/ 10.1177/1362168810383329

Awwad, A., & Tavakoli, P. (2022). Task complexity, language proficiency and working memory: Interaction effects on second language speech performance. *International Review of Applied Linguistics in Language Teaching*, *60*(2), 169-196. https://doi.org/10.1515/iral-2018-0378

Bassano, D., & van Geert, P. (2007). Modeling continuity and discontinuity in utterance length: a quantitative approach to changes, transitions and intra-individual variability in early grammatical development. *Developmental Science*, *10*(5), 588-612. https://doi.org/10.1111/j.1467-7687.2007.00629.x

Bulté, B., & Housen, A. (2012). Defining and operationalizing L2 complexity. In A. Housen, F. Kuiken, & I. Vedder. (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). John Benjamins.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23-48). Longman.

Cameron, L., & Larsen-Freeman, D. (2007). Preview article: Complex systems and applied linguistics. *International Journal of Applied Linguistics, 17*(2), 226-240.

de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition, 10*(1), 7-21. https://doi.org/10.1017/S1366728906002732

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Ellis, R. (2008). *The study of second language acquisition* (2nd edition). Oxford University Press.

Ellis, R., & Barkhuizen G. (2005). *Analyzing learner language.* Oxford University Press.

Evans, R., & Larsen-Freeman, D. (2020). Bifurcations and the emergence of L2 syntactic structures in a complex dynamic system. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.574603

Ferrari, S. (2012). A longitudinal study of complexity, accuracy and fluency variation in second language development. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity accuracy and fluency in SLA* (pp. 277-298). John Benjamins.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354-375. https://doi.org/ 10.1093/applin/21.3.354

Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics*, *36*, 98-116. https://doi.org/10.1017/S0267190515000082

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 1-20). John Benjamins.

Jackson, D. O., & Suethanapornkul, S. (2013). The cognition hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning, 63*(2), 330-367.

Kormos, J. (2006). *Speech production and second language acquisition.* Lawrence Erlbaum.

Lambert, C., & Kormos, J. (2014). Complexity, accuracy, and fluency in task-based L2 research: Toward more developmentally based measures of second language acquisition. *Applied Linguistics*, *35*(5), 607-614. https://doi.org/10.1093/applin/amu047

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics, 18*(2), 141-165. https://doi.org/10.1093/applin/18.2.141

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590-619. https://doi.org/10.1093/applin/aml029

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

Larsen-Freeman, D. (2018). Looking ahead: Future directions in, and future research into, second language acquisition. *Foreign Language Annals, 51*(1), 55-72. https://doi.org/10.1111/flan.12314

Larsen-Freeman, D. (2020). Complexity theory: Relational systems in interaction and in interlocutor differences in second language development. In L. Gurzynski-Weiss (Ed.), *Cross-theoretical explorations of interlocutors and their individual differences* (pp. 189-208). John Benjamins.

Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford University Press.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), 387-417.

Levkina, M., & Gilabert, R. (2012). The effects of cognitive task complexity on L2 oral production. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity accuracy and fluency in SLA* (pp. 171-198). John Benjamins.

Li, C., & Zhang, L. J. (2023). The development of accuracy and fluency in second language (L2) speaking related to self-efficacy through online scaffolding:

A latent growth curve modeling analysis. *Journal of Psycholinguistic Research, 52*(5), 1371-1395. https://doi.org/10.1007/s10936-023-09950-7

Lowie, W. (2017). Lost in state space? Methodological considerations in complex dynamic theory approaches to second language development research. In L. Ortega & Z. Han (Eds), *Complexity theory and language development: In celebration of Diane Larsen Freeman* (pp. 123-141). John Benjamins.

Lowie, W., & Verspoor, M. (2019). Individual differences and the ergodicity problem. *Language Learning, 69*, 184-206. https://doi.org/10.1111/lang.12324

Lowie, W., & Verspoor, M. (2022). A complex dynamic systems theory perspective on speaking in second language development. In T. M. Derwing, M. J. Munro, & R. I. Thomson (Eds.), *The Routledge handbook of second language acquisition and speaking* (pp. 39-53). Routledge.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Lawrence Erlbaum.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing, 15*(3), 323-338. https://doi.org/10.1093/llc/15.3.323

Michel, M. C. (2017). Complexity, accuracy and fluency in L2 production. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 50-68). Taylor & Francis.

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

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. https://doi.org/10.1093/applin/amp044

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

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

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27-57. https://doi.org/10.1093/applin/22.1.27

Robinson, P. (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. John Benjamins.

Skehan, P. (1996). Second language acquisition research and task-based instruction. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 17-30). Heinemann.

Skehan, P. (1998). *A cognitive approach to language learning.* ERIC.

Skehan, P. (2001). Tasks and language performance assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 167-185). Longman.

Skehan, P. (2014). Limited attentional capacity, second language performance, and task-based pedagogy. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 211-260). John Benjamins.

Skehan, P., Bei, X., Li, Q., & Wang, Z. (2012). The task is not enough: Processing approaches to task-based performance. *Language Teaching Research*, *16*(2), 170-187. https://doi.org/10.1177/1362168811428414

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*(3), 185-211. https://doi.org/10.1177/136216889700100302

Taguchi, N. (2013). Production of routines in L2 English: Effect of proficiency and study-abroad experience. *System, 41*(1), 109-121. https://doi.org/10.1016/j.system.2013.01.003

Tonkyn, A. P. (2012). Measuring and perceiving changes in oral complexity, accuracy and fluency: Examining instructed learners' short-term gains. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity accuracy and fluency in SLA* (pp. 221-244). John Benjamins.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, *17*(1), 84-119. https://doi.org/10.1093/applin/17.1.84

van Geert, P. (1994). *Dynamic systems of development: Change between complexity and chaos*. Harvester Wheatsheaf.

van Geert, P. (1995). Growth dynamics in development. In R. F. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 313-337). The MIT Press.

van Geert, P., & Steenbeek, H. (2005). Explaining after by before: Basic aspects of a dynamic systems approach to the study of development. *Developmental Review*, *25*(3-4), 408-442. https://doi.org/10.1016/j.dr.2005.10.003

van Geert, P., & van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior & Development, 25*(4), 340-374. https://doi.org/10.1016/S0163-6383(02)00140-6

Vercellotti, M. L. (2017). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistics, 38*(1), 90-111. https://doi.org/10.1093/applin/amv002

Verspoor, M., & de Bot, K. (2022). Measures of variability in transitional phases in second language development. *International Review of Applied Linguistics in Language Teaching*, *1*, 85-101. https://doi.org/10.1515/iral-2021-0026

Verspoor, M., Lowie, W., & de Bot, K. (2021). Variability as normal as apple pie. *Linguistics Vanguard*, *7*(S2). 20200034. https://doi.org/10.1515/lingvan-2020-0034

Verspoor, M., & van Dijk, M. (2011). Visualizing interaction between variables. In M. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 85-98). John Benjamins.

Yu, H., & Lowie, W. (2020). Dynamic paths of complexity and accuracy in second language speech: A longitudinal case study of Chinese learners. *Applied Linguistics*, *6*, 855-877. https://doi.org/10.1093/applin/amz040

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