



Development of linguistic complexity in repeated form-focused classroom tasks: A longitudinal study of Japanese university EFL learners

SAEKO TOYOSHIMA  

Tsuda University, Japan

Abstract

This longitudinal classroom-based study examined how linguistic complexity developed in a Japanese university English as a Foreign Language (EFL) course that integrated task repetition with embedded focus on form. While previous research has emphasised group-level trends in complexity development, few studies have explored how learners develop along different trajectories in authentic classroom settings. This study addressed two questions: (1) how does multidimensional linguistic complexity change across repeated form-focused tasks, and (2) what learner profiles emerge from this development?

Written texts were collected from 35 beginner-level students (CEFR A1–A2) across eight task cycles over the academic year. Lexical diversity, lexical density, and mean sentence length were analysed, and k-means cluster analysis was used to identify learner profiles. The results showed gradual but non-linear development. Rather than progressing uniformly, learners followed three distinct developmental patterns that reflected different combinations of lexical diversification and syntactic elaboration. These findings suggest that repeated performance tasks with embedded form-focused support create conditions for structural development, while allowing differentiated developmental pathways to emerge. A profile-based perspective may help teachers better understand learner variation and support complexity development in classroom contexts.

Keywords: Linguistic complexity, focus on form approach, performance task, Japanese EFL learners

Introduction

The development of linguistic complexity remains one of the central concerns in second language acquisition (SLA) because it reflects learners' ability to deploy increasingly elaborate lexical and syntactic resources during communication (Bulté & Housen, 2012; Housen & Kuiken, 2009; Norris & Ortega, 2009). Within the widely used complexity–accuracy–fluency (CAF) framework (Housen & Kuiken, 2009; Skehan, 2009), complexity refers to the range and elaboration of linguistic resources used by learners. However, complexity is not a unitary construct. Lexical diversity, lexical density (LexD), and sentence length may develop in different ways, and improvement in one dimension does not automatically lead to growth in another (Housen et al., 2019; Norris & Ortega, 2009). This multidimensional perspective suggests that classroom development may be uneven rather than strictly linear.

Another related issue is how development is examined. Several studies have reported group averages over short instructional periods. While useful, such analyses can mask differences among learners. From a dynamic perspective, variability is a natural feature of language development (Larsen-Freeman, 2006). Even when students receive the same instruction, they may progress in different ways, depending on their starting level, learning strategies, and responsiveness to feedback. However, longitudinal classroom-based research exploring learner-level differences, especially among beginner English as a Foreign Language (EFL) students, is limited.

Task-based language teaching (TBLT) offers an approach that supports complexity development. Task repetition has been shown to reduce cognitive load and allow learners to refine their language use (Bygate, 2001; DeKeyser, 2018). When learners repeat similar communicative tasks, they may shift their attention from meaning to form and gradually expand their linguistic resources. In addition, focus on form (FonF), or paying brief attention to linguistic features within communicative activities (Ellis, 2016; Long, 1991), may help learners notice and incorporate new structures.

Most previous studies on repetition have examined oral tasks in controlled research settings (Bygate, 2001; Foster & Skehan, 1996; Révész, 2009; Yuan & Ellis, 2003). Writing plays a central role in many EFL classrooms at universities. Written tasks allow for more planning and revision than oral production (Ortega, 1999), which may influence how complexity develops over time. However, most previous studies have adopted short-term experimental or cross-sectional designs, typically examining one or two repetitions over a limited instructional period. Consequently, little is known about how repeated written tasks shape complexity development across an extended classroom curriculum.

This issue is particularly relevant in the context of Japanese university EFL. English education has traditionally emphasised grammar and examination preparation; however, many institutions are now incorporating communicative and task-based approaches. Repeated performance tasks and digital portfolio assignments were recently introduced in a Japanese university EFL course by the present author, who suggested the effectiveness of students' engagement in communicative activities in their target language (Toyoshima, 2024). However, there is limited empirical evidence showing how such practices influence structural development over time, especially among beginner learners (CEFR A1–A2) whose linguistic resources are still emerging. In the context of Japanese university EFL, recent applied linguistics research has also highlighted the impact of broader educational shifts, including the expansion of English-medium instruction and communicative pedagogies (Rose & McKinley, 2018).

In light of these considerations, the present study investigates how multidimensional linguistic complexity develops across repeated classroom tasks with embedded FonF. Rather than assuming uniform progress, this study adopts a learner-level perspective to examine how

different dimensions of complexity evolve over time. In addition to examining overall classroom trends, cluster analysis was used to identify learner profiles that reflect differentiated developmental trajectories. By longitudinally examining the development of complexity in an authentic Japanese university EFL context, this study aims to contribute to a more nuanced understanding of how instructional practices interact with learner variability to shape structural development.

Literature Review

Conceptualising Linguistic Complexity in SLA

The development of linguistic complexity has been a central concern in the study of SLA. Within the widely used CAF framework (Housen & Kuiken, 2009; Skehan, 2009), complexity refers to the elaboration and diversification of linguistic resources in learner production. However, complexity is not a single construct. It encompasses a variety of dimensions, including lexical diversity, lexical sophistication, clausal elaboration, and sentence length (Bulté & Housen, 2012; Norris & Ortega, 2009).

Recent research has emphasised that lexical and syntactic complexity may evolve semi-independently (Housen et al., 2019). Learners may increase sentence length while relying on familiar vocabulary or by diversifying vocabulary within relatively simple syntactic frames. These findings challenge conventional linear models of development and underscore the multidimensional nature of complexity.

Furthermore, interpretation of complexity measures requires a cautious approach. As indicated by Ortega (1999), indices such as Type-Token Ratio (TTR) are sensitive to text length, whereas mean sentence length (MSL) may reflect discourse organisation rather than purely syntactic advancement. For this reason, TTR findings are often interpreted in conjunction with other complexity indices rather than as standalone measures (Norris & Ortega, 2009). Consequently, researchers advocate the concurrent examination of multiple indices and the interpretation of findings within the confines of contextual constraints (Norris & Ortega, 2009).

Another issue that merits attention is the level of analysis employed. Many studies have reported group-level averages across time, which may obscure learner-level variability (Larsen-Freeman, 2009). According to Larsen-Freeman (2006), the notion of variability in SLA is not merely random noise but rather an inherent feature of development. However, a paucity of classroom-based studies has been observed in the modelling of differentiated developmental trajectories, particularly in beginner EFL contexts.

Task Repetition and Complexity Development

According to Robinson (2011), TBLT places significant emphasis on communicative interactions as the primary catalyst for acquisition. In contrast to the presentation-practice-production (PPP) approach (Ellis, 2020), TBLT integrates meaningful communication with opportunities for attention to form. Within the TBLT framework, task repetition is a pedagogical technique that has garnered significant attention.

Task repetition has been demonstrated to enhance processing efficiency and promote proceduralisation (Bygate, 2001; DeKeyser, 2018). When learners repeat tasks, reduced cognitive load may allow for greater allocation of attentional resources to linguistic form and structural elaboration. Yuan and Ellis (2003) demonstrated that planning time and repetition positively influence complexity and fluency. Hu (2018) and Kobayashi and Kobayashi (2018) found that recurrent involvement in tasks fosters lexical and syntactic refinement.

However, these findings are not consistent. Several studies have indicated that an increase in fluency is accompanied by only marginal gains in syntactic complexity (Foster &

Skehan, 1996). This suggests the presence of trade-offs among performance dimensions. The impact of repetition may be contingent on factors such as task design, planning conditions, and learner proficiency (Révész, 2009). Moreover, a significant proportion of extant research has focused on oral production. The scope of existing longitudinal research examining the repeated administration of written tasks in beginner university EFL contexts remains limited.

Writing is a distinct mode of expression that enables more extensive planning and revision opportunities than oral production (Ortega, 1999). Consequently, the implementation of repeated written tasks in a semester-long curriculum may yield cumulative developmental effects that differ from those observed in short-term experimental designs. However, there is a paucity of research on how repetition shapes multidimensional complexity trajectories over extended instructional sequences in classroom settings.

FonF in Classroom-Based TBLT

FonF, originally conceptualised by Long (1991), refers to brief, contextually embedded attention to linguistic elements that arise during meaning-focused communication. In contrast to conventional PPP approaches, which compartmentalise grammatical instruction from communicative tasks, FonF integrates form-based attention into communicative tasks (Ellis, 2016).

The theoretical underpinnings of FonF are rooted in the noticing hypothesis (Schmidt, 1990) and interactionist frameworks of SLA. It has been demonstrated that a deliberate focus on linguistic features during communicative performance can facilitate the restructuring of interlanguage systems, particularly when learners are developmentally prepared. Research conducted in classroom settings has demonstrated that form-focused feedback and guided attention can enhance grammatical development and structural awareness (Ellis, 2001; Samuda, 2001).

In the context of beginner EFL instruction, the incorporation of FonF in repeated tasks may hold particular significance. Low-proficiency learners frequently employ simple sentence frames and high-frequency vocabulary. A systematic approach to linguistic forms in meaningful tasks has been shown to gradually expand lexical and syntactic resources without disrupting communicative engagement. However, it should be noted that learners may differ in their ability to internalise such input. Lexical enrichment and syntactic elaboration are two possible priorities.

Despite the proliferation of research on TBLT and FonF, empirical studies examining the combined effects of task repetition and embedded form-focused instruction on multidimensional written complexity remain scarce, particularly in Japanese university settings. Most extant studies have either examined short-term experimental interventions or focused on a single performance dimension. From a pedagogical perspective, applied linguistics research has increasingly emphasised the need to align classroom practices with real-world language use and communicative demands (Murray, 2020).

A Synopsis of the Research Gap and its Rationale

In summary, previous studies indicate that linguistic complexity is multidimensional and non-linear, task repetition can facilitate proceduralisation and structural refinement, and FonF supports the integration of meaning and form in communicative contexts. However, there is a paucity of research on how repeated form-focused tasks influence learner-specific complexity trajectories across extended instructional sequences in beginner EFL classrooms.

Despite the growing adoption of communicative and task-based approaches in Japanese university English education, longitudinal evidence documenting how beginner EFL learners develop linguistic complexity remains limited. Most classroom-based studies have been conducted in other educational contexts or have focused on short-term interventions, leaving

the developmental trajectories of Japanese university learners underexplored. This context is particularly important because many Japanese university students have experienced extensive grammar-focused instruction but comparatively limited opportunities for sustained communicative language use. Examining complexity development within this setting therefore contributes not only to classroom-based SLA research but also to a better understanding of how task-based instruction may support linguistic development in contexts undergoing pedagogical transition.

Accordingly, this study addresses the following research questions:

- RQ1: What is the process through which multidimensional linguistic complexity develops across repeated tasks with embedded FonF in beginner university EFL classrooms?
- RQ2: What learner profiles emerge from this development at the learner level?

Methods

Research Design and Context

This study employed a longitudinal classroom-based design to examine the development of lexical and syntactic complexity across repeated performance tasks. The study was conducted in a required English course at a Japanese university. The pedagogical approach of the course entailed a task-based instructional model with an integrated FonF component. This task format was replicated eight times over the course of one academic year.

The instructional design aimed to balance communicative meaning-making with systematic attention to the target linguistic forms introduced in each textbook unit. Rather than adopting the conventional approach of isolating grammar instruction, form-focused feedback was integrated into the performance tasks. This design provides an ecologically valid setting in which complexity development can be observed under authentic classroom conditions.

Participants

The participants were 35 first-year university students enrolled in a compulsory English course. Their proficiency levels ranged from CEFR A1 to A2, as determined by the placement procedures administered by the institution. The students had completed six years of English education in secondary school; however, their experience with extended communicative, oral, and written tasks was limited.

All students provided consent for the use of their written data for research purposes. The study was conducted in accordance with institutional ethical guidelines. To safeguard participant confidentiality, all personally identifiable information was removed from the dataset prior to analysis.

Task Design and Instructional Procedure

The focal task was a performance-based digital portfolio task introduced previously (Toyoshima, 2024). In each unit, the students were tasked with creating presentation slides related to a theme aligned with the textbook. These themes included a wide range of personal and professional subjects, including favourite music, dream jobs, and achievements. The target grammatical form for each unit (e.g., simple present, comparative forms, and modal verbs) was introduced through instructional input and brief guided practice activities prior to task performance (see Appendix A).

Subsequently, students created digital slides that incorporated the target forms and relevant vocabulary (example slides shown in Appendix B). These portfolio presentations were conducted during designated ‘portfolio sessions’ held four times a year, resulting in eight assignment cycles throughout the academic year. Following each session, peer evaluation was conducted using a rubric that assessed content, visual design, and the appropriateness of English usage (see Appendix C). The instructor provided brief feedback on form.

This iterative task cycle enabled learners to re-examine similar communicative demands while progressively assimilating novel linguistic forms. In this case, the repetition was cumulative rather than identical, as topics varied across units while maintaining a consistent performance structure.

Data Collection

The dataset consisted of written texts extracted from students’ digital presentation slides. Across the eight tasks, 248 texts were collected, with minor variations due to missing data. The length of the texts also varied across tasks, a phenomenon attributable to the divergent scope of topics and the level of learner output required by each task.

It should be noted that the analysis was conducted on the final version of each task submitted for assessment. This study did not analyse spoken delivery during presentations. Emphasis was placed on the development of written production to ensure the comparability of results across tasks and to mitigate variability associated with spontaneous oral performance.

Data Preparation and Text Processing

To ensure consistency, all text was converted into a plain-text format prior to analysis. Tokenisation was implemented consistently across the entire dataset. The term ‘word’ was defined as an orthographic unit separated by spaces. Contractions (e.g., ‘I’m’) were counted as two tokens (‘I’ + ‘am’), and numerals were excluded from lexical calculations. The retention of proper nouns was contingent on their function as lexical items in the text.

Subsequent sentence segmentation was guided by punctuation-based criteria. A sentence was defined as a unit ending with a full stop, question mark, or exclamation mark. In instances where punctuation was absent, segmentation was determined based on syntactic completeness, defined as the presence of a subject–predicate structure. These procedures were applied consistently across all texts.

To minimise distortion caused by length variation, lexical diversity was interpreted with caution alongside LexD and MSL. The objective of this study was to establish comparability rather than to propose novel complexity indices. The linguistic measures used in this study (TTR, LexD, and MSL) were generated automatically using predetermined computational procedures that were applied consistently across the entire corpus. Text preprocessing, including tokenisation and sentence segmentation, followed explicit operational criteria described above to ensure procedural consistency. Because the measures were derived algorithmically rather than through subjective qualitative coding, formal inter-rater reliability was not applicable to the calculation of the linguistic indices.

Measures of Linguistic Complexity

To this end, three indices were selected to capture complementary dimensions of complexity:

Type-Token Ratio (TTR):

The TTR was calculated as the number of unique word types divided by the total number of tokens in each text (Ellis, 2009; Housen & Kuiken, 2009). The TTR is widely used as an

indicator of lexical diversity in classroom-based SLA. However, it is sensitive to text length (Malvern et al., 2004; Tweedie & Baayen, 1998).

In the present study, the TTR was interpreted cautiously for three reasons. First, the texts were produced within a relatively constrained task format (digital presentation slides), which limited extreme variation in length. Second, the TTR was interpreted alongside mean token counts and other indices (LexD and MSL), allowing potential length-related distortions to be contextualised rather than interpreted in isolation. Third, given the beginner proficiency level of the participants (CEFR A1–A2) and the pedagogical orientation of the study, the TTR provides an accessible and widely interpretable measure of lexical diversity in classroom contexts (Norris & Ortega, 2009).

Nevertheless, findings related to TTR should be interpreted with caution, as mean token counts differed substantially across tasks (e.g., Task 2: $M = 80.6$; Task 5: $M = 136.6$).

Lexical Density (LexD):

LexD is calculated as the proportion of lexical words (nouns, verbs, adjectives, and adverbs) relative to the total number of tokens (Dahl, 2004). This measure reflects the informational load of a text and has been widely used to examine the extent to which learners rely on content words rather than function words in their production.

As LexD is less sensitive to overall text length than lexical diversity measures, it provides a useful complementary perspective on lexical development. In classroom-based writing, higher LexD may indicate an increased ability to express ideas using more information-rich lexical items.

Mean Sentence Length (MSL):

MSL was calculated by dividing the total number of tokens by the number of sentences in each text (Ortega, 1999). Although MSL does not directly measure clausal subordination, it is frequently used as a global indicator of syntactic elaboration in learner production.

In beginner-level writing, in which complex clause structures may occur infrequently, sentence-level measures provide a practical way to capture broad tendencies toward structural expansion. Longer sentences may reflect an increasing use of coordination, subordination, and elaborated noun phrases as learners attempt to express more complex ideas.

The selection of these indices was based on their widespread use in classroom-based SLA research and their interpretability within pedagogical contexts.

Cluster Analysis Procedure

To identify distinct learner profiles, k-means clustering was conducted using Python (version 3.11) and the scikit-learn library (Pedregosa et al., 2011). Prior to clustering, each learner's mean TTR, mean LexD, and mean MSL were calculated by averaging the scores across all tasks in which they participated. These three aggregated values were then standardised to z-scores to ensure comparability across indices with different numerical scales.

The optimal number of clusters (k) was determined by examining the within-cluster sum of squares across k values ranging from 2 to 6. The resulting plot showed an elbow-type pattern at $k = 3$, indicating that additional clusters produced diminishing reductions in within-cluster variance. Therefore, a three-cluster solution was selected as the most parsimonious and interpretable model.

Clustering was conducted at the learner rather than task level. Each learner was represented by aggregated complexity indices across tasks, allowing the analysis to capture overall developmental tendencies rather than task-specific fluctuations. The aim of cluster analysis was

descriptive rather than predictive, providing a profile-based perspective on learner variability within a shared instructional environment. All computational analyses were conducted by the author using predefined analytical procedures implemented consistently across the dataset. Interpretation of the findings was undertaken by the author within the theoretical framework of the study.

Results

To address RQ1, overall task-level trends in lexical and syntactic complexity were first examined using the mean indices across the eight task cycles. Subsequently, to address RQ2, a k-means cluster analysis was conducted to identify differentiated learner profiles based on aggregated complexity measures.

Results for RQ1: Task-Level Complexity Development

Table 1 shows the task coverage and overall trends for Tasks 1 to 8.

Table 1. *Task-level descriptive statistics for complexity indices*

Task	Files	Mean tokens	Mean TTR	Mean LexD	Mean MSL
1	32	106.3	0.659	0.632	12.22
2	29	80.6	0.686	0.587	15.3
3	32	110.2	0.584	0.562	13.92
4	32	110.3	0.611	0.537	11.06
5	31	136.6	0.616	0.529	9.59
6	31	135.8	0.609	0.601	11.43
7	30	138.1	0.577	0.527	12.85
8	31	111.1	0.604	0.573	12.19

The results indicated that Task 2 exhibited the highest mean TTR (0.686), while Task 7 demonstrated the lowest (0.577). Regarding LexD, Task 1 exhibited the highest mean value (0.632), whereas Task 7 demonstrated the lowest (0.527). Mean MSL was highest for Task 2 (15.30) and lowest for Task 5 (9.59). Overall, TTR values fluctuated across tasks rather than showing a consistent linear increase, suggesting that lexical diversity varied depending on task topic and communicative demands. Conversely, LexD and MSL decreased. Rather than interpreting these shifts as regressions, they may reflect topic effects and changes in how learners package information. As students’ confidence increases, they may engage in more efficient writing, contingent on the communicative objective of the topic.

Despite the observed fluctuations, several patterns merit closer interpretation. The highest mean TTR was observed for Task 2, whereas Task 7 demonstrated comparatively lower lexical diversity and LexD. Rather than interpreting these fluctuations as simple regressions, it is important to consider the influence of topic demands and communicative purposes. The nature of the tasks was such that some required the presentation of narratives (e.g., achievements in the past), whereas others encouraged the use of evaluative or persuasive discourse. Such variations may have influenced the manner in which learners allocated their linguistic resources.

Across tasks, mean token counts generally increased from the middle stages of the course, suggesting a greater willingness to elaborate. However, these increases in text length do not uniformly correspond to increases in lexical diversity or density. This divergence underscores the multidimensional nature of complexity development. In some instances, learners produced longer texts using familiar lexical items, whereas in others, shorter texts demonstrated greater lexical concentration.

Results for RQ2: Learner Profile Differentiation

Cluster analysis revealed three distinct developmental profiles (Table 2). To better understand the developmental implications of these clusters, it is useful to compare initial and later task performances.

Table 2. *Cluster profiles of linguistic complexity*

Cluster	n	Mean TTR	Mean LexD	Mean MSL
Cluster 1	12	0.640	0.571	11.29
Cluster 2	19	0.590	0.559	11.11
Cluster 3	4	0.619	0.591	18.69
Total	35	—	—	—

Cluster 1 (n = 12) exhibited gradual lexical diversification over time. These learners exhibited relatively rich lexical use, accompanied by gradual increases in lexical diversity and moderate syntactic complexity. Initial texts were characterised by brevity and repetition. However, subsequent tasks demonstrated an increase in the use of abstract vocabulary and the expansion of explanatory structures. This suggests a progressive shift from personal description toward more elaborated discourse.

The largest group, designated Cluster 2 (n = 19), exhibited relatively stable lexical patterns while demonstrating a gradual increase in structural elaboration. The resulting texts are indicative of fluent yet cautious writing, characterised by the stable use of familiar vocabulary and relatively simple sentence structures. The learners appeared to prioritise clarity and fluency during the study. Their development was marked by a gradual and incremental process rather than a sudden or dramatic transformation.

Syntactic elaboration, as indicated by sentence length and structural complexity, was relatively high in Cluster 3 (n = 4) during the early stages of the study. The presence of longer sentences and embedded clauses was also observed in Task 1. However, LexD did not increase proportionally, indicating that syntactic advancement does not necessarily imply broader lexical diversification.

Figure 1 illustrates these differences across clusters. As shown in Figure 1, Cluster 3 was characterised by a substantially higher MSL, whereas Cluster 1 showed slightly higher lexical diversity, as indicated by TTR.

These patterns indicate trade-offs: some learners expand sentence length and elaboration, whereas others diversify their vocabulary within shorter texts. This phenomenon underscores the limitations of a single index for capturing developmental nuances across learners. Cluster 3 comprises a relatively small group of learners. These learners appear to demonstrate higher

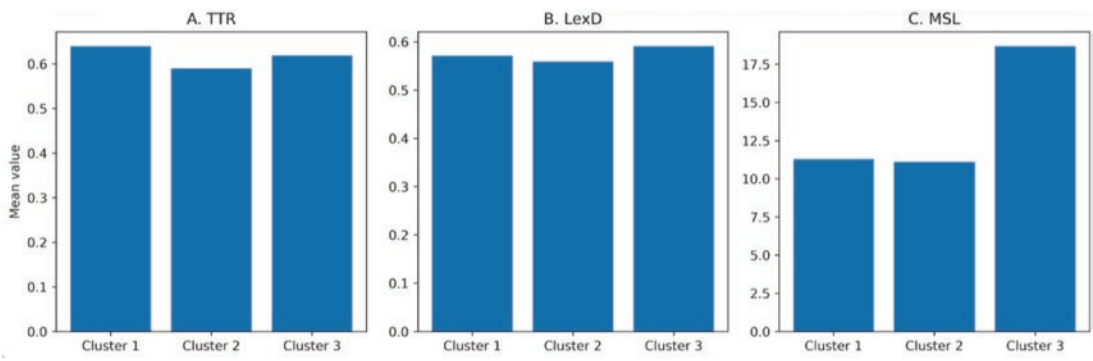


Figure 1. Cluster comparison of complexity indices

Note: Panel A shows mean Type–Token Ratio (TTR), Panel B shows mean lexical density (LexD), and Panel C shows mean sentence length (MSL) for each learner cluster.

levels of syntactic elaboration from an early stage, distinguishing them from most students who show more gradual developmental patterns.

Representative Student Excerpts

To illustrate the characteristics of the three learner profiles identified in the cluster analysis, representative excerpts from three students are presented below. The students were referred to as Student A (Cluster 1), Student B (Cluster 2), and Student C (Cluster 3). For each student, excerpts were selected from Tasks 1 (early stage), 4 (mid stage), and 8 (late stage) to illustrate changes across the instructional sequence. Each excerpt represents a single text produced by the learner, and the associated TTR, LexD, and MSL values are reported for that text.

Cluster 1: Gradual lexical diversification (Student A)

Student A represented learners in Cluster 1 who showed gradual lexical diversification accompanied by moderate syntactic expansion.

Task 1: Excerpt (TTR = 0.69, LexD = 0.69, MSL = 4.36)

My favourite music is pop music because it makes me happy. I often listen to music when I am studying or relaxing. Music is important for my life.

In the initial stage, the text consisted of short sentences with repeated vocabulary (e.g., music), reflecting limited lexical diversity and simple sentence structures typical of beginner-level production.

Task 4: Excerpt (TTR = 0.66, LexD = 0.49, MSL = 9.56)

There are many people who have special talents, but I think a person who can continue 努力 [effort] is really talented. For example, athletes practice every day to improve their skills and never give up. A talented person is not only someone who is famous but also someone who works hard and supports others.

Compared with Task 1, the student employed a wider range of lexical items (e.g., talent, athletes, improve, and supports). The sentences also became longer, with the use of coordination and explanatory clauses.

Task 8: Excerpt (TTR = 0.67, LexD = 0.54, MSL = 16.20)

If my city builds a skateboard park, young people will have a safe place to practice and communicate with others. This will make the city more active and help create a stronger community.

In the final stage, the student used conditional structures and more abstract vocabulary (e.g., community, communicate), indicating a shift toward more elaborated discourse.

Cluster 2: Stable lexical use with gradual structural expansion (Student B)

Student B belonged to Cluster 2, which was the largest group in the dataset. These learners tended to maintain relatively stable lexical repertoires while gradually expanding their sentence structures.

Task 1: Excerpt (TTR = 0.64, LexD = 0.59, MSL = 9.13)

My favourite music is pop music because it is easy to listen to and makes me feel happy.

The sentence structure was simple but already somewhat longer than that of Cluster 1 learners, reflecting moderate syntactic elaboration.

Task 4: Excerpt (TTR = 0.62, LexD = 0.53, MSL = 15.13)

My dream for the future is to get a job that I enjoy and live a stable life. To achieve this dream, I need to study hard.

The text introduces future-oriented expressions and purpose structures while maintaining a relatively familiar lexical repertoire.

Task 8: Excerpt (TTR = 0.59, LexD = 0.56, MSL = 10.33)

If the city holds more music events, young people can enjoy music together and the city will become more lively.

This stage shows the use of conditional structures and slightly more abstract discourse, although lexical diversity remains relatively stable.

Cluster 3: Early syntactic elaboration (Student C)

Student C represented Cluster 3, a smaller group characterised by comparatively advanced syntactic elaboration from an early stage.

Task 1: Excerpt (TTR = 0.73, LexD = 0.61, MSL = 25.00)

My favourite movie is an action movie because it has exciting scenes and characters that make people feel strong emotions.

Even in the initial task, the learner produced relatively long sentences containing embedded clauses.

Task 4: Excerpt (TTR = 0.67, LexD = 0.62, MSL = 12.00)

A person with great talent is someone who can understand others and build better relationships in society.

The text demonstrates continued use of abstract vocabulary and relational concepts such as society and relationships.

Task 8: Excerpt (TTR = 1.00, LexD = 0.63, MSL = 16.00)

If my city creates more public spaces, people can communicate and feel connected to the community.

A high TTR value reflects the short length of the excerpt rather than unusually high lexical diversity.

These representative excerpts illustrate how learners within each cluster followed different developmental pathways:

Cluster 1: Learners gradually expanded lexical diversity and discourse elaboration.

Cluster 2: Learners maintained relatively stable lexical patterns while slowly increasing their structural complexity.

Cluster 3: Learners demonstrated comparatively advanced syntactic elaboration in the early stages of the course.

Therefore, the excerpts provide qualitative support for the quantitative patterns identified through cluster analysis, illustrating how multidimensional complexity develops differently across learners within the same instructional environment.

These comparisons show that learners did not reach a common developmental endpoint. Instead, the data reveal the emergence of multiple patterns within a single instructional environment. The presence of distinct profiles underscores the importance of examining learner-level trajectories rather than relying solely on aggregate classroom means. In addition to qualitative distinctions among clusters, relative changes across tasks further elucidate developmental tendencies. A comparison between Tasks 1 and 8 indicates that learners in Cluster 1 exhibited the largest relative increase in lexical diversity, suggesting a gradual expansion of vocabulary resources over time. Conversely, Cluster 2 demonstrated subtle changes in lexical indices accompanied by gradual increases in sentence length, indicating incremental structural elaboration. Cluster 3 exhibited a consistent pattern of comparatively higher MSLs from the outset. This finding suggests that syntactic resources were already relatively well developed at the initial stage.

Of particular significance is the observation that these differences are not only absolute but also proportional. Although mean values fluctuated depending on task topic, the relative positioning of the clusters remained stable across the sequence. This stability lends further support to the interpretation that differing developmental patterns are not merely coincidental but reflect systematic processes. Thus, cluster-based comparison supports the view that repeated task engagement does not result in uniform convergence; rather, it facilitates the emergence of multiple trajectories of complexity development within a single instructional context.

Discussion

Responses to the Research Questions

The findings of this study indicate that multidimensional linguistic complexity does not develop in a uniformly linear manner across repeated classroom tasks. Although overall classroom trends suggest gradual development, cluster analysis revealed three distinct learner profiles,

demonstrating that learners followed differentiated developmental trajectories within a shared instructional environment.

From a developmental perspective, these findings suggest that complexity development may be better understood as the redistribution of linguistic resources rather than simple accumulation. Instead of improving uniformly across all dimensions, learners appeared to selectively allocate attentional and cognitive resources. This interpretation aligns with multidimensional and limited-capacity models of second language performance, which propose that learners cannot simultaneously maximise all aspects of linguistic complexity because of attentional constraints (Housen et al., 2019; Norris & Ortega, 2009; Skehan, 2009). Consequently, some learners prioritise lexical diversification, whereas others extend syntactic structures within relatively stable lexical repertoires. These differentiated patterns reflect adaptive developmental processes shaped by both cognitive constraints and instructional experience.

From a multidimensional complexity perspective, the results support prior research suggesting that lexical and syntactic dimensions may evolve semi-independently (Housen et al., 2019; Norris & Ortega, 2009). For example, learners in Cluster 1 demonstrated gradual lexical diversification and moderate structural expansion, whereas learners in Cluster 2 maintained relatively stable lexical repertoires while increasing their structural elaboration. Despite its smaller size, Cluster 3 exhibited comparatively advanced syntactic elaboration from an early stage. These findings reinforce the view that growth in one dimension does not necessarily entail parallel development in others, thus reflecting the interaction of partially independent linguistic subsystems.

Importantly, the findings indicate that group-level averages may obscure meaningful learner-level variation. While overall classroom trends suggested gradual development, cluster analysis revealed stable and differentiated learner profiles. This observation is consistent with dynamic perspectives in second language acquisition, which emphasise that variability, rather than random fluctuation, is an inherent feature of development (Larsen-Freeman, 2006, 2009). From this perspective, divergence among learners reflects differences in prior knowledge, attentional focus, and responsiveness to instructional input.

The design of the present study integrated task repetition with an embedded FonF, and the observed developmental patterns were broadly consistent with theoretical predictions derived from each component. Task repetition is theorised to reduce cognitive load, potentially freeing attentional resources for linguistic forms (Bygate, 2001; DeKeyser, 2018), while brief form-focused attention within communicative tasks promotes noticing and restructuring (Ellis, 2002, 2016). However, it is important to note that this study could not isolate the contributions of either component. The observed complexity development may reflect the combined instructional design, general language exposure over the academic year, maturation effects, or learner-internal factors independent of instruction. Without a control group or experimental manipulation of individual components, causal attribution to task repetition or FonF is not warranted. Therefore, the present findings are best understood as evidence that complexity development occurred within this instructional environment rather than as evidence for the efficacy of any particular instructional mechanism.

Furthermore, the persistence of distinct learner profiles suggests that instructional affordances interact with internal learner factors to shape developmental trajectories. Although repeated exposure and feedback create opportunities for restructuring, learners engage with them based on their developmental readiness and attentional priorities (Ellis, 2002; Shintani, 2018). Therefore, examining development through a profile-based lens provides a more nuanced understanding of how complexity evolves in classroom contexts.

Taken together, these findings suggest that complexity development in classroom-based EFL settings is multidimensional, variable, and learner-specific. Rather than progressing toward a single uniform endpoint, learners appear to follow individualised developmental trajectories shaped by both cognitive constraints and instructional conditions. This perspective underscores the value of learner-level analysis in understanding second language development in authentic classrooms. This interpretation is also consistent with classroom-based applied linguistics research highlighting the importance of interaction, learner engagement, and individual developmental trajectories in language learning (Sato & Ballinger, 2016).

Pedagogical Implications

Several implications emerge for classroom implementation. First, educators who employ repeated performance tasks should recognise that learners do not necessarily demonstrate uniform progress. The development of language skills can be observed in multiple ways. For some students, this development is evident in the diversification of their vocabulary. In contrast, others demonstrate a shift toward syntactic elaboration, referring to increased sentence complexity and the use of more sophisticated grammatical structures. Recognition of these varied pathways may encourage instructors to adopt more flexible evaluation criteria, moving away from reliance on a single indicator of ‘improvement’. These findings, based on the patterns identified in RQ2, underscore the importance of incorporating learner-level analysis into classroom research.

In addition, the integration of FonF within repeated tasks represents a promising pedagogical approach. Rather than isolating grammar instruction from communicative activities, an alternative strategy involves integrating form awareness into meaningful performance tasks. This pedagogical approach has the potential to facilitate gradual qualitative transformation in language learning. In beginner EFL instruction, such integration may be particularly valuable, as learners frequently rely on limited sentence frames and high-frequency vocabulary. The provision of systematic yet concise feedback can promote incremental restructuring without disrupting communicative flow.

Furthermore, the findings have implications for assessment design. If complexity development is multidimensional, assessment practices should incorporate multiple indicators rather than rely on a single holistic measure. The portfolio-based tasks employed in the present study may facilitate tracking of longitudinal development across multiple dimensions. Teachers may benefit from monitoring lexical and syntactic growth separately to better understand individual learners’ progress.

Finally, the variability observed across clusters underscores the importance of differentiated instructional support. Some learners may require targeted scaffolding to expand their syntactic resources, whereas others may benefit from lexical enrichment strategies. Recognition of learner profiles may enable instructors to provide feedback that aligns more closely with individual developmental needs.

Curriculum and Contextual Implications

In the context of Japanese university EFL instruction, which is characterised by large classes and exam-oriented traditions, repeated performance-based tasks with embedded FonF may present an alternative approach for achieving gradual complexity development. The longitudinal structure of the course enabled learners to re-examine communicative demands while progressively assimilating new forms. This cumulative design may be particularly well suited to novice learners who require repeated exposure and opportunities for consolidation.

Concurrently, the modest size of Cluster 3 indicates that advanced syntactic elaboration may emerge only among a subset of learners in beginner cohorts. Consequently, curriculum planners should anticipate heterogeneity rather than uniform outcomes. The provision of varied opportunities for expansion, such as optional extension tasks or guided elaboration prompts, has the potential to support learners at different developmental stages.

Limitations and Directions for Future Research

It is important to acknowledge the limitations of this study. First, the sample size was relatively small, particularly in Cluster 3 ($n = 4$). Although the three-cluster solution is interpretable and pedagogically meaningful, the stability of these clusters cannot be assumed to extend beyond the present dataset. To enhance the generalisability of the findings, replication with larger cohorts across multiple institutions is necessary. Future studies may also compare alternative clustering methods (e.g., hierarchical clustering or model-based approaches) to examine the robustness of learner profile identification.

Second, the participants in the present study were beginner-level learners (CEFR A1–A2). Although this proficiency range reflects the characteristics of the target instructional context, developmental trajectories may differ among intermediate- or advanced-level learners, who possess broader lexical and syntactic resources and may therefore demonstrate different patterns of complexity development. Future research should examine whether similar learner profiles emerge across different proficiency levels and educational contexts.

Third, a restricted set of complexity indices was employed. The following variables were analysed: TTR, LexD, and MSL. Although these measures are widely used in classroom-based SLA research and are pedagogically interpretable, they do not encompass all aspects of linguistic complexity. Clause-level complexity and lexical sophistication measures were intentionally not included because the learner texts consisted of relatively short digital presentation scripts produced by beginner-level learners (CEFR A1–A2). Under these conditions, sentence-level and lexical measures provide more stable and pedagogically interpretable indicators of development, while more fine-grained indices may be less reliable and more difficult to interpret. For example, TTR is sensitive to text length, whereas MSL does not necessarily reflect clausal subordination or syntactic sophistication in a fine-grained manner. Despite the use of multiple indices to reduce reliance on a single metric, future research may consider incorporating additional indices, such as clause-level complexity or lexical sophistication measures, to provide a more comprehensive account of the developmental process.

A related limitation concerns the use of TTR as a measure of lexical diversity. Because TTR is sensitive to text length, comparisons across tasks with varying token counts may partially reflect length effects rather than genuine lexical development. Although TTR was interpreted alongside LexD and MSL to mitigate this limitation, more robust length-independent indices such as MTLT (McCarthy & Jarvis, 2010) or vocd-D (Malvern et al., 2004) could provide more reliable estimates. These indices were not employed in the present study due to constraints related to data processing and the pedagogical orientation of the analysis; however, their inclusion would strengthen future research.

Fourth, this study focused exclusively on written production derived from digital presentation slides. Although written tasks allow for planning and revision, the findings cannot be directly generalised to oral performance. The relationship between the development of written and spoken complexity remains an important area for future research. A more comprehensive understanding of multidimensional development can be achieved by examining whether similar learner profiles emerge in oral tasks.

The fifth limitation concerns the causal interpretability of the findings. This study employed a single-group longitudinal design without a comparison or control condition. Consequently, it is not possible to attribute the observed complexity development to any specific instructional component, such as task repetition, embedded FonF, general language exposure, or their interaction. The observed patterns are consistent with theoretical predictions from TBLT and FonF literature; however, they do not constitute evidence that these mechanisms were causally responsible for the observed development. Future research employing experimental or quasi-experimental designs with control groups is required to isolate the effects of specific instructional components. Similarly, the dual role of the present researcher as both course instructor and data analyst introduces potential observer bias in task implementation and interpretation, which further limits causal claims.

Although the linguistic indices were calculated automatically using predefined procedures, the preprocessing of learner texts (e.g., sentence segmentation in cases where punctuation was absent) was conducted by the researcher according to explicit criteria. While these procedures were applied consistently across the dataset, future studies could further strengthen methodological transparency by asking an independent researcher to verify preprocessing decisions for a subset of the corpus before automated analysis.

The dual role of the teacher as both instructor and researcher may introduce potential bias in task implementation and interpretation. Despite efforts to maintain consistency and transparency in data processing, future research conducted across multiple instructors or institutions may enhance ecological validity and reduce the influence of context-specific factors.

Despite these limitations, this study provides insights into differentiated developmental trajectories in an authentic instructional setting. Addressing these limitations may contribute to a more nuanced understanding of how complexity develops during repeated engagement with form-focused tasks among beginner EFL learners.

Conclusion

This study examined the development of linguistic complexity across repeated form-focused tasks in a Japanese university EFL classroom. Using a longitudinal and learner-level approach, the study identified three distinct developmental profiles, showing that learners do not progress uniformly even within a shared instructional environment.

The findings suggest that repeated performance tasks with embedded focus on form can create conditions for gradual structural development in beginner classrooms. At the same time, the results highlight the importance of adopting a profile-based perspective when interpreting learner development.

Overall, this study contributes to classroom-based SLA research by showing how longitudinal task engagement can reveal differentiated developmental pathways that may not be visible in group-level analyses.

Acknowledgement

I am grateful to the students whose task outcomes contributed to this study. They appeared to enjoy learning English through the tasks, despite feeling nervous about studying English. This classroom practice has also provided a foundation for the next stage of my research.

References

- Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In Housen, A., Kuiken, F., & Vedder, I. (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.
- Dahl, Ö. (2004). *The growth and maintenance of linguistic complexity*. John Benjamins. <https://doi.org/10.1075/slcs.71>
- DeKeyser, R. (2018). Task repetition for language learning: A perspective from skill acquisition theory. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 27–41). John Benjamins.
- Ellis, R. (2001). Introduction: Investigating form-focused instruction. *Language Learning*, 51, 1–46. <https://doi.org/10.1111/j.1467-1770.2001.tb00013.x>
- Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? Review of the research. *Studies in Second Language Acquisition*, 24(2), 223–236. <https://doi.org/10.1017/S0272263102002073>
- Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, 30(4), 474–509. <https://doi.org/10.1093/applin/amp042>
- Ellis, R. (2016). Focus-on-form: A critical review. *Language Teaching Research*, 20(3), 405–428. <https://doi.org/10.1177/1362168816628627>
- Ellis, R. (2020). Teacher-preparation for task-based language teaching. In C. Lambert & R. Oliver (Eds.), *Using tasks in second language teaching: Practice in diverse contexts* (pp. 99–119). Multilingual Matters.
- 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. <https://doi.org/10.1017/S0272263100015047>
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461–473. <https://doi.org/10.1093/applin/amp048>
- Housen, A., De Clercq, B., Kuiken, F., & Vedder, I. (2019). Multiple approaches to complexity in second language research. *Second Language Research*, 35(1), 3–21. <https://doi.org/10.1177/0267658318809765>
- Hu, X. (2018). Effects of task type, task-type repetition, and performance criteria on L2 oral production. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 143–169). John Benjamins.
- Kobayashi, E., & Kobayashi, M. (2018). Second language learning through repeated engagement in a poster presentation task. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 223–254). John Benjamins.
- 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>
- Long, M. (1991). Focus-on-form: A design feature in language teaching methodology. In DeBot, K., Ginsberg, R., & Kramsch, C., *Foreign language research in cross-cultural perspective* (pp. 39–52). John Benjamins.
- Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development*. Palgrave Macmillan.
- McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392. <https://doi.org/10.3758/brm.42.2.381>
- Murray, N. (2020). English as a lingua franca and pedagogical implications. *Australian Journal of Applied Linguistics*, 3(2), 117–135. <https://doi.org/10.29140/ajal.v3n2.273>
- 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>
- Ortega, L. (1999). Planning and focus-on-form in L2 oral performance. *Studies in Second Language Acquisition*, 21, 109–148. <https://doi.org/10.1017/S0272263199001047>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., &

- Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <http://jmlr.org/papers/v12/pedregosa11a.html>
- Révész, A. (2009). Task complexity, focus-on-form and second language development. *Studies in Second Language Acquisition*, 31(3), 437–470. <https://doi.org/10.1017/S0272263109090366>
- Robinson, P. (2011). Task-based language learning: A review of issues. *Language Learning*, 61, 1–36. <https://doi.org/10.1111/j.1467-9922.2010.00624.x>
- Rose, H., & McKinley, J. (2018). Japan's English-medium instruction initiatives and the globalization of higher education. *Australian Journal of Applied Linguistics*, 1(1), 1–14. <https://doi.org/10.29140/ajal.v1n1.3>
- Samuda, V. (2001). Guiding relationships between form and meaning during task performance: The role of the teacher. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 119–140). Longman.
- Sato, M., & Ballinger, S. (Eds.). (2016). *Peer interaction and second language learning: Pedagogical potential and research agenda*. John Benjamins. <https://doi.org/10.1075/llt.45>
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Shintani, N. (2018). Mediating input-based tasks for beginner learners through task repetition. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 255–278). John Benjamins.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510–532. <https://doi.org/10.1093/applin/amp047>
- Toyoshima, S. (2024). Performance tasks as a mediator to communicating in English and for enhancing language learner engagement. *SAGE Open*, April-June, 1–13. <https://doi.org/10.1177/21582440241240544>
- Tweedie, F. J., & Baayen, R. H. (1998). How variable may a constant be? *Computers and the Humanities*, 32, 323–352. <https://doi.org/10.1023/A:1001749303137>
- 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>

Appendices

Appendix A

Topics and Targeted Forms of Each Task

Task 1: Listen to My Favourite Music/Let's Watch My Favourite Movie

Form: Simple present tense

Task 2: This is My 'Non-negotiable' Thing

Form: Simple present with adverbs of frequency

Task 3: How about this 'Dream' Job?

Form: 'like' vs. 'would like'

Task 4: Someone with Great Abilities

Form: can/can't

Task 5: Can I Introduce My Best Friend?

Form: Modifying adverbs

Task 6: The Wildest Place in My Hometown

Form: Comparative and superlative adjectives

Task 7: 'I did it!': An Achievement in My Life

Form: Simple past tense

Task 8: I Want ... near My Campus! (each student filled in the blank)

Form: should/shouldn't

Appendix B

Example Student Presentation Slides

Task 3: How about This ‘Dream’ Job?

I have a dream job ...

A **teacher**!

WHEN I MET MY DREAM JOB...



I like school. And I like teacher.
Because They helped me when I was
having trouble with study. They always
kindly teach me how to study.
Also, They are always kind and listens to
my concerns. It helped my heart.



IF I GET THE DREAM JOB...



I want kindly teach children to study.
And I want to help children as I was.
I want to create an environment where
children can enjoy school.
I want to have fun with children every day.

IF I GET THE DREAM JOB...

I want to help children study and support
children.
I want to take children out to school who
don't go to school.
I want to lead children into a future that they
will not regret as adults.



Appendix C
Peer Evaluation Rubric

Presenter's Name									
Contents (Match with the topic and interesting idea)	5	3	1	5	3	1	5	3	1
Visuals (Design & Appealing)	5	3	1	5	3	1	5	3	1
English (Appropriateness)	5	3	1	5	3	1	5	3	1
Comment									