Journal of Second Language Writing 54 (2021) 100856

Contents lists available at ScienceDirect

Journal of Second Language Writing

journal homepage: www.elsevier.com/locate/jslw

Emergence of multiple groups of learners with different writing-development trajectories in classroom: Growth mixture modeling

Kyoko Baba [a,*], Ryo Nitta [b]

[a] Department of English, Kinjo Gakuin University, 2-1723 Omori, Moriyama-ku, Nagoya, 463-8521, Japan
[b] Rikkyo University, Center for Foreign Language Education and Research, 3-34-1 Nishi-Ikebukuro, Toshima-ku, Tokyo, 171-8501, Japan

ARTICLE INFO

Keywords:
Multiple learner groups
Classroom
Writing development
Complex dynamic systems theory
Growth mixture model

1. Introduction

When tasks are investigated in terms of variables (e.g., task types, task sequence, and task-implementation variables), the effect of a variable is examined under the assumption that learners under each condition (level) comprise a statistically homogeneous population. Yet, this assumption may not always be met since, as many teachers would agree, every learner is different and how they perform a task may vary from learner to learner. Different tasks may produce different effects, but if a task elicits different performance on the part of learners, what they learn from the task will also vary. In addition, when the task is repeatedly performed, it is possible that learners will orient and act on a task differently each time, resulting in diverging trajectories of language development. If so, there can be learners who greatly benefit from engaging in a task multiple times on one hand and those who learn less from it on the other. Thus, it is of significance to investigate the difference between them and seek ways to scaffold those who learn less from it.

This study is part of a larger research project that explores inter-learner differences in development of L2 writing proficiency in an intact classroom setting. The goal of the project is to find clues for identifying factors that positively influence learners' developmental

* Corresponding author.
E-mail addresses: kbaba@kinjo-u.ac.jp (K. Baba), ryo.nitta@rikkyo.ac.jp (R. Nitta).

https://doi.org/10.1016/j.jslw.2021.100856
Received 8 May 2021; Received in revised form 27 October 2021; Accepted 28 October 2021
Available online 1 December 2021

trajectories. To achieve this goal, this project takes a two-step approach: (a) after a writing task procedure was repeatedly used in the real classroom context throughout one academic year, learners' developmental patterns are analyzed, and (b) if any inter-learner difference exists, groups of learners with a different developmental pattern are compared to determine factors that might have affected the difference, which may suggest ways of scaffolding various types of learners. The purpose of this study was primarily the first step, that is, to examine whether there are multiple developmental patterns in writing proficiency of more than 100 L2 learners. If there were differences, possible factors affecting the differences were also explored.

1.1. Task iteration and L2 development in classroom

Variability has not been the central concern in task based language teaching (TBLT) research; however, it is informative and a valid object of study in complex dynamic systems theory (CDST) research. The CDST perspective questions the identical and fixed effect of a task, and takes into account that variability will arise when L2 learners engage in a task (a task is broadly defined as including what may be considered an activity from TBLT). The variability in task performance will lead to inter-individual variability in development. To highlight variability in task performance, CDST distinguishes *repetition* and *iteration* of tasks, which are not interchangeable. Repetition implies that learners repeat the same task in the same way, while iteration suggests that "each time the same task is used, the learners' experience of it will be different, in part because learners will orient to it differently" (Larsen-Freeman, 2018, p. 317). That is, when a learner performs a specific task multiple times, they will perform it in a different way every time because the previous experience with the task may change the way the learner perceives, understands, orients, and acts on it the next time they perform it. This also means that individual students in a class perform a task differently, and what they learn from the iteration varies from learner to learner. Through this iterated process, learners act as active agents and pursue their own unique developmental trajectories.

When the repeated engagement with a task is conceptualized as iteration instead of repetition, it is not operationalized as a variable because it is impossible to treat it in a research design as a factor that is supposed to cause a homogeneous effect in every learner. Rather, it functions as a parameter to which individual learners respond differently. The concept of iteration, coupled with other CDST notions, such as non-linearity and dynamism, have led researchers to investigate inter-individual differences in developmental trajectories in terms of various aspects of oral and written L2 performance: oral and written proficiency (Baba & Nitta, 2014; Baba, 2020; Huang, Steinkrauss, & Verspoor, 2021; Larsen-Freeman, 2006; Lowie & Verspoor, 2019), linguistic complexity in oral and written discourse (Bulté & Housen, 2018; Chan, Verspoor, & Vahtrick, 2015; Vyatkina, Hirschmann, & Golcher, 2015; Wang & Tao, 2020), self-regulation and other learner-psychology factors in writing (Han & Hiver, 2018; Jackson & Park, 2020; Nitta & Baba, 2018), and authorial voice (Fogal, 2019, 2020a). For example, Chan et al. (2015) showed notable differences in the developmental processes of L2 oral and written syntactic complexity even in a pair of identical twins. Similarly, Lowie and Verspoor (2019) investigated the development of writing proficiency of a homogeneous group of secondary-school students and found that their developmental trajectories diverged, even after controlling for the effect of motivation and aptitude.

CDST studies emphasizing such individual differences in development may give an impression that every learner follows a completely unique developmental path, which cannot be compared and generalized to other learners. This may hold true at an individual level, for no two learners follow the same developmental trajectory, which is dynamic, non-linear, and interactive with other systems/sub-systems (i.e., adaptive). Yet, if these idiosyncratic individual trajectories are analyzed at a group level, it may be possible to find some "nomothetic knowledge about idiographic [developmental] processes" in dynamic and complex systems (Molenaar, 2015, p. 40), and if so, it would provide situated and practical pedagogical implication for teachers in classrooms. However, many CDST studies investigating L2 writing (except for some studies such as Baba (2020), Huang et al. (2021), and Lowie and Verspoor (2019)) have been conducted on an individual level focusing on small cases; therefore, more research is needed to uncover inter-individual differences in L2 writing development at a group level.

1.2. Modeling multiple developmental trajectories

On approach to investigate idiosyncratic individual trajectories at a group level is called *person-centered analyses* (e.g., Howard & Hoffman, 2017; Jung & Wickrama, 2008; Nagin, 2005). A person-centered analysis aims "to group individuals into categories, each one of which contains individuals who are similar to each other and different from individuals in other categories" (Muthén & Muthén, 2000, p. 882). For example, a growth mixture model (GMM) is used to classify individuals who share similar developmental patterns into groups (e.g., Muthén, 2004; Muthén & Muthén, 2017).

A GMM deals with longitudinal (or multi-wave) data, as does individual growth curve (IGC) modeling, but the treatment of variables is different. IGC modeling allows variations between individual trajectories, and if individuals can be classified into groups based on a certain variable(s) (e.g., task, task condition), it is possible to assess the effect of the variable(s). In other words, "existence of distinct developmental trajectories must be assumed a priori" (Nagin, 2005, p. 11) on the assumption that students with a specific property or condition will follow different developmental trajectories, that is, variables are the principal concern in IGC modeling.

In contrast, a GMM places importance on classifying individuals into groups in accordance with the shape of trajectories, in which "individuals are treated in a more holistic fashion [than variable-centered research] by focusing on a system of variables taken in combination rather than in isolation" (Meyer & Morin, 2016, p. 584). Groups of people with different trajectories are called *latent trajectory classes*, because they are not manifest a priori. With a GMM, it is assumed that latent trajectory classes may come from a different population, as opposed to the IGC modeling assumption that all the individuals come from one population. After exploring latent trajectory classes, it is also possible to add covariates, that is, predictors (antecedents) and distal outcomes (consequences) of class membership (Wickrama, Lee, O'Neal, & Lorenz, 2016). Predictors are possible factors that might have affected the categorization of classes, and distal outcomes are possible subsequent outcomes that different trajectory classes will achieve. The addition of covariates is considered a combination of person-centered and variable-centered approaches, and this extended model is called the conditional GMM or general growth mixture model (GGMM) (Muthén & Muthén, 2000).

In summary, a GMM is used to examine whether there are multiple groups of people with similar developmental patterns, and a GGMM is used to determine factors that explain the grouping. In this study, we hoped to find answers to the question of what makes learners trace diverged developmental trajectories by using these models.

1.3. Current study

Baba (2020) used IGC modeling without covariates to investigate how individual students' trajectories of L2 writing changed over one year and revealed that on the whole, the quality of their writing increased, but the rate of change varied significantly from person to person. She then chose sample high-growth and low-growth students (five students in each group; ten in total) based on the slope of their trajectories and qualitatively explored the L2 compositions and reflective comments in the L1 of the two groups. As a result, the high-growth performers tended to write constructive and future-directed reflective comments, analyze their own performances logically and in detail, and devise and adopt various measures and strategies to develop their writing abilities. In addition, the high-growth performers made agentic effort to improve their writing performance for the task by using various means available to them. In contrast, low performers wrote brief and monotonous reflective comments on how they wrote a composition instead of constructive and future-directed ones. This may explain why they failed to reflect on and amend problematic behaviors: they tended to write compositions too cautiously even though they were instructed to concentrate on writing as much as possible for this specific task (e.g., paying too much attention to word selection and grammar) and wrote superficial compositions without elaborating a theme. It is then reasonable to assume that these differences in attitude and the way they perform the task at one time will affect subsequent engagements with the task.

This study thus explored whether L2 learners show divergent trajectories of L2 writing development at a group level through task iteration. If this was the case, characteristics of each group of learners with a similar developmental pattern were explored in terms of two distal outcomes (final text length and final course grade). We then assessed the effect of two predictors (type of reflection and degree of variability). The covariates were supposed to explain the learners' developmental endpoints (distal outcomes) and contributors to diverged developmental trajectories (predictors) in a complemental manner. The selection of the covariates, however, had to be exploratory in nature due to the lack of empirical research and based on previous findings and practical considerations. The first distal outcome is the final text length, which was the variable of focus for this study (see the data and indices section). The final course grade is added as an indicator of how each group of students performed in the class in which the writing task was iterated. The type of reflective comments (either short or extended, see the methods section) is chosen as a predictor because Baba (2020) suggested that it may contribute to divergent developmental patterns. Degrees of variability is the factor that previous research has shown to be associated with the larger gain in writing proficiency (Bulté & Housen, 2018; Huang et al., 2021; Lowie & Verspoor, 2019), so it is included as a predictor. Thus, the research questions of the current study were as follows:

1 Do multiple groups with a distinct developmental trajectory of text length emerge when L2 learners iterate the same task procedure every week over one academic year? If so, do the groups perform differently in terms of final text length and final course grade?
2 Is group membership predicted by the type of reflective comments and degree of variability?

2. Method

2.1. Setting

The data-collection site was a mid-sized private women's university located in the central area of Japan. The writing data were collected over a year from five different student cohorts (that is, the data were collected five times) attending the same university course that one of the authors taught over one academic year (30 weeks in total). As the core features of the course content remained basically the same (e.g., the same textbooks, and the similar handouts with some modifications in wording were used), the students learned similar content with the same syllabus.

The course was one of the mandatory EFL courses for first-year students and was the only course that focused on developing academic writing skills. All the students belonged to the English department of the university and were majoring in English literature, culture, or linguistics. The class met once a week for 90 min. The class usually involved a narrative timed-writing task, which was the focus in the current study, a translation activity from Japanese to English focusing on grammar, and activities for learning academic writing such as paragraphing and logical argumentation. The students were required to write two academic essays per semester, and the focused timed-writing tasks were the only opportunity for them to do narrative-like writing. Participation in the study was voluntary (those who consented to participate signed a consent form), but all the students engaged in the timed-writing task as part of the course.

2.2. Participants

The participants of the study were 105 first-year female students (the university was a women's university). The students' mother tongue was Japanese, and they had received six years of formal English education before entering university. All the students took the

TOEIC Listening and Reading test at the beginning (May) of the academic year, and their average score was 392.60 ($SD = 59.93$). Note that all the participants took the TOEIC test in the same format. As described below, the type of reflective comments (short or extended) is one of the predictors in the GGMM, and only the first cohort of students wrote short reflective comments. There was no significant difference in initial L2 proficiency between the first and other cohorts: the average TOEIC Listening and Reading scores were 390.43 ($SD = 37.93$) for the first cohort and 393.21 ($SD = 65.02$) for the other cohorts, $t(62.14) = -.26$, $p = n.s.$

Before starting university, most students had never composed a timed narrative-writing task like the one used in the study, either in English or in Japanese. The typical writing activity in their junior high school and secondary school studies was to translate Japanese sentences into English, so this focused task was the first experience for most of them to write their own ideas and thoughts.

2.3. Task and procedure

The task used in this study was timed narrative writing, in which students wrote a composition on a chosen topic for 10 min (see narrative_writing_form_short_Japanese, narrative_writing_form_short_English, narrative_writing_form_extended_Japanese and narrative_writing_form_extended_English, available in the online version of this article. The students in this study used the Japanese forms, and the English forms are the translations by the authors). Before the first composition, they were given detailed instructions on how to perform the writing task. Students were told that the main purpose of the task was for them to become more comfortable writing in English and that they should not pay too much attention to grammatical errors and instead write as much as possible. The students were supposed to strengthen various aspects of writing proficiency in the course, and it was necessary to make them aware that the purpose of the narrative writing task was different from that of the academic essay writing task that was also practiced. This point was emphasized repeatedly throughout the academic year. Each time, students were given a list with three different topics to choose from. The aim of offering three choices was to compensate for differences in students' individual experiences and preferences (see topiclist, available in the online version of this article). Most topics were personal, and no topic required significant background knowledge. The topics were written both in Japanese and English, and the same list of three topics was used for two successive weeks. Students were told to write on the same chosen topic twice. A new topic list was then handed to them the following week. The same set of topic lists was used in the same order for all five cohorts, with only a few minor revisions.

Immediately after they had written their compositions, the students wrote reflective comments on their writing in Japanese (all comments cited in this work were translations by the authors). The students in the first cohort wrote brief personal remarks in a small box under their composition (usually a few sentences) (see narrative_writing_form_short_Japanese(English)). The box did not specify what they should write about, but typical remarks that they wrote about were concerning their weaknesses. Hereafter, this type of reflection is referred to as short reflection.

A major revision to the form for the reflective comments was made in the second year, and the revised form was used from then on (see narrative_writing_form_extended_Japanese(English)). That is, the students in the second through fifth cohorts wrote longer comments regarding more specific aspects of writing by using the same revised template. This type of reflection is called extended reflection. The revised form was double-sided (the narrative-writing task was on one side, and the extended reflection was on the other side), providing students with more space to write extensive reflective comments. In addition, the extended reflection side provided six boxes specifying a topic for comments. The first four boxes facilitated reflection on various aspects of the day's writing: (a) grammar and vocabulary, (b) organization and expression, (c) content (such as episodes and examples), and (d) writing processes and strategies. Although the main purpose of the narrative writing task was to familiarize the students with writing in English, they were expected to expand their focus to other aspects of writing whenever possible, and the four boxes provided such an opportunity. The next box was a space for comparing the day's composition with previous ones (the previous composition was at hand when they filled out this box but not during the composition phase). In the last box, they wrote goals for future compositions (e.g., in what way they would like to make progress). The students were told to complete all six boxes. The reflection form usually took about 10 min to complete.

The teacher wrote a few encouraging comments on every composition and reflection every week and returned them the following week. As the aim of the feedback was to create a sense of audience and to maintain students' motivation to write every week, no error was corrected. The feedback was returned after students had finished writing their next compositions, so they did not have access to their previous compositions or the feedback while composing the next compositions.

2.4. Data and indices

A total of 2947 compositions were collected. The narratives were typed out and processed through a software application that provided the word count for each composition. All spelling mistakes in the compositions were corrected manually and with a spell checker.

The index that represents individual composition is text length, which was operationalized as the total number of words in a composition written in a specific time limit (10 min in this case). This type of index may usually be considered to denote fluency of writing and roughly indicates how quickly one can write within the time limit. Yet, it is neither the best index for writing fluency (e.g., Van Waes & Leijten, 2015) nor the sole index of writing quality. Writing proficiency should be viewed holistically, so other aspects of writing and the relationship among different aspects as well as their relationship to higher-level factors are also important (Byrnes, 2020; Fogal, 2020b). Thus, the current study simply regarded it as a unit of analysis, or a representative index, because the emphasis was placed not on the analysis of writing fluency or quality in the general sense, but on the investigation of developmental patterns of one specific variable. It was then necessary to choose an optimized window of analysis that is appropriate, sensitive to change, and practical/feasible. First, the index seemed appropriate because the main objective of this specific narrative writing task was to let the students write as much as possible for them to get used to writing in English, which was explicitly and repeatedly told to the students. Therefore, focusing on changes in text lengths stood to reason. In addition, when it comes to timed writing tasks, the length (i.e., how much one can write within the time limit) is a notable characteristic shared by essays that are evaluated highly (Friginal, Li, & Weigle, 2014; Jarvis, Grant, Bikowski, & Ferris, 2003). Regarding sensitivity to change, an optimal index should detect changes in compositions, that is, it needs to appropriately discriminate compositions written by the same writer at different times. Text length seemed sufficient because Baba and Nitta (2014), using the same timed narrative-writing task, demonstrated that it was able to capture changes in basic writers' writing. Lastly, the index of text length was practical and robust. There was a substantive number of compositions to be analyzed, so it was unfeasible to use variables that include subjective judgements. Note, however, many other CAF (complexity accuracy fluency) measures enable similar non-subjective analyses.

As mentioned in the introduction section, the current study examined the effect of two distal outcomes (final text length and final course grade) and two predictors (type of reflection and degree of variability). The final text length is represented by the average of the last five compositions because the number of words in only the last composition may be exceptional (e.g., some student may have been in a bad mood that day and could not write as much as usual). The final course grade was standardized within each cohort. For all cohorts, the teacher evaluated the performance of each student consistently on the basis of the same criteria and syllabus, and gave a score (ranging from 0 to 100) at the end of the academic year. The score reflected the results of the mid-term (20 %) and final examinations (25 %) (each of which included narrative timed writing), two academic essays (45 %), and participation (10 %). The two academic essays were evaluated holistically in consideration of five aspects (i.e., the appropriateness of the task and topic, organization, coherence, use of language, and appeal to the reader). As such, the evaluation criteria was the same across the five cohorts, but each cohort consisted of different students, so z scores standardized within each cohort, instead of raw scores, were used for subsequent analyses. Regarding the predictors, type of reflection is a categorial variable, and takes two values, either short or extended. The degree of variability is represented as the coefficient of variation (CV), which is calculated for each student by dividing the standard deviation by the average text length of all the compositions that the student wrote. The CV shows the dispersion of a specific variable without regard to measurement unit.

2.5. Analysis

A GMM was used to explore distinct groups of developmental trajectories (latent trajectory classes). GMM is least sensitive to collinearity, and appropriate for repeated measures (Tu, Tilling, Sterne, & Gilthorpe, 2013). After deciding the number of latent trajectory classes, the conditional GMM (i.e., GGMM) was used. All analyses were conducted with Mplus Version 8.4. Mplus is a statistical package for structural equation modeling and can be used for a wide range of statistical analyses as well as their combinations with a simple command language. It is particularly strong with multilevel analyses (analyses at both individual and inter-individual levels), including the GMM and GGMM.

There is no absolute criteria for the sample size in using a GMM. Meyer and Morin (2016) recommend more than 300, but the small sample size of the current study ($N = 105$) is compensated for with relatively large measurement points (about 30). Diallo and Morin (2014) have shown that if there are 10 measurement points, the minimum number of participants should be 50, ideally 100 is needed; thus, the data of the current study satisfied this criterion.

We followed the analytical procedures described by Wickrama et al. (2016). According to Wickrama et al. (2016), there are three main estimation problems in a GMM: problems related to a non-normal distribution, local maxima, and convergence. The normality of the data was checked by carefully examining information criteria statistics, specifically, the Lo-Mendell-Rubin likelihood ratio test (LMR–LRT), Bayesian information criterion (BIC), and the sample size adjusted Bayesian information criterion (SSABIC), which have been shown to be accurate in identifying the number of latent classes (Nylund, Asparouhov, & Muthén, 2007). As a result, the LMR–LRT, BIC, and SSABIC values were satisfactory for the optimal model (see the results section). Local maxima problems occur when the model-estimation process is prematurely stopped due to sub-optimal solutions and the global solution is not reached. Our optimal model turned to have a global solution. A GMM often produces the convergence problem due to complicated modeling, but our model did converge.

The selection of the optimal unconditional GMM is exploratory in nature, which means that it is impossible to absolutely judge the extent to which the data fit the model with goodness-of-fit indices such as the standardized root mean residual (SRMR), root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis index (TLI) (Meyer & Morin, 2016). Three points were considered in the selection of an optimal model: statistical indices, the number and size of classes, and interpretability of class trajectories. First, the statistical indices used in the current study were BIC, SSABIC, the adjusted LMR–ALRT (Adj. LMR–LRT), bootstrapped likelihood ratio test (BLRT), and entropy values. There is no cut-off value for the fit indices, but smaller values are better. Entropy indicates classification accuracy and takes values between 0 and 1. Entropy values of .80 or higher suggest high class separation.

The second point to be considered is the number and size of latent trajectory classes. Although a higher number of classes tends to result in a better model, a more parsimonious model (i.e., fewer classes) should be selected. In addition, more classes may include classes with a small number of members. Yet, such a model with a small class is allowed "if the researcher can accurately defend what is gained from this small class given the possibility of low power and a lack of statistical precision" (Wickrama et al., 2016, p. 215).

The third point is whether latent class trajectories are interpretable. A model with many latent class trajectories may result in better values on the statistical indices. However, this makes the interpretation of each trajectory difficult because it is likely that some have a similar pattern, making it difficult to distinguish their characteristics. In addition, statistical results must not be adopted unquestionably. They should be corroborated with theory and interpreted so that the results are meaningful in the literature with pedagogical implications.

In the current study, four latent models were compared: latent class growth analysis (LCGA), the GMM, the GMM with fixed intercepts, and the GMM with fixed slopes. LCGA is a restricted version of a GMM, in which within-class variations are fixed to zero. That is, it is assumed that intercepts and slopes may vary across different latent trajectory classes but that individuals within each class have the same intercept and slope. GMMs relax this assumption and allows variations across and within latent trajectory classes. GMMs either with fixed intercepts or slopes are partially constrained GMMs, in which either intercepts or slopes are fixed within each class. The latent models were estimated for each of the two, three, and four latent trajectory classes, resulting in 12 latent models.

Before selecting the number of latent trajectory classes, traditional latent growth curve models with a linear and quadratic form were estimated to choose a functional form for individual trajectories (the basic shape of the resulting trajectories). Although the values for the quadratic curve were slightly better than the linear curve, the linear curve was selected because the interpretation of the results is more straightforward for the linear curve (Singer & Willett, 2003). The purpose is to choose an approximate form of representation, so the choice of the linear curve is not to argue that developmental trajectories are linear.

After deciding on an optimal GMM, covariates (predictors and distal outcomes) are included into the model (GGMM). There are two ways to estimate GGMMs: the direct specification approach and the 3-step approach (Wickrama et al., 2016). In the former, covariates are directly included in a GMM, and the latter estimates the effect of the covariate after selecting an unconditional GMM. Thus, in the direct specification approach, the class formation is affected by the addition of covariates to this approach, while it is protected in the 3-step approach. The current study adopted the 3-step approach because the focus was more on the exploration of latent trajectory classes than the assessment of the effect of covariates; thus, it was important to protect the class formation estimated using the optimal unconditional GMM. The 3-step approach can be executed either manually or with the AUXILIARY option of Mplus (Asparouhov & Muthén, 2013). In the current study, the estimation of GGMMs were estimated with both methods. However, their results were approximately the same (as observed by Wickrama et al., 2016), so only those with the AUXILIARY option are presented in the results section.

3. Results

The results are presented in the order of the research questions. To answer Research Question 1, a traditional growth curve model was estimated prior to the latent models to determine whether the data were likely to come from a single population; in which case, GMMs with the assumption of heterogeneous trajectory classes are inappropriate. To make this judgement, two criteria were used: model-fit indices and variances of growth parameters. First, the model-fit indices of the growth curve model were χ^2 (df) = 1406.13 (460), $p < .001$; RMSEA (90 %) = 0.14 (.132, .148); CFI/TLI = 0.75/0.76; SRMR = .11. The cut-off values are .06 or below for RMSEA, .95 or greater for CFI/TLI, and 0.08 or below for SRMR (Wickrama et al., 2016). Therefore, the growth curve model poorly fit the data, meaning that the single-population model was not appropriate for the current dataset. Second, the variances of the growth parameters (i.e., the intercept and slope parameters, because a linear model is selected, as explained above) were statistically significant: 81.57, $p < .001$ for the intercept parameter, and 0.57, $p < .001$ for the slope parameter. These results indicate that both the initial text length (the intercept) and its change rate (the slope) significantly differed among the students. That is, there were individual differences in how many words the students wrote within 10 min at the beginning and to what extent they made progress in writing more words, which suggests that heterogeneous developmental trajectories may exist.

Next, as described in the analysis section, four latent models (LCGA, the GMM, GMM with fixed intercepts, and GMM with fixed slopes) for two, three, four classes were estimated (see model_summaries that shows the results of all the models, which is available in the online version of this article). For LCGA, Adj. LMR–LRT did not produce any significant *p*-values (.21, .18, and .07 for two, three, and four classes respectively). BIC and SSABIC values were also larger than those for the other models. Thus, LCGA models were left out.

Regarding the GMMs, only the two-class model produced significant *p*-values for Adj. LMR–LRT (.04) and BLRT (.00). However, the members were unevenly distributed between the two classes in the two-class model: Class 1 contained most of the participants (102), and Class 2 contained only three. The results may suggest that the three individuals had substantially different developmental trajectories from the rest of the students, but it has little to say about the trajectories of the majority of the students. Hence, the two-class GMM was also discarded.

GMMs with either fixed intercepts or slopes produced better results than the above models. The two- and four-class GMMs with fixed intercepts reached significant *p*-values for both Adj. LMR–LRT (.001 for the two-class model and .01 for the four-class model) and BLRT (.001 for the two-class model and .00 for the four-class model). For both models, the entropy values were satisfactory (.94 and .93 respectively), and the distribution of class membership looked fair although Class 4 of the four-class model had only three members. These two models were kept for further comparison.

With regard to GMMs with fixed slopes, the two-class and three-class models produced significant *p*-values for both Adj. LMR–LRT (.00 for the two-class model and .001 for the three-class model) and BLRT (.00 for both models). However, the two-class model had the same problem with the two-class GMM, that is, Class 1 contained the majority of participants, while Class 2 contains only three. Therefore, the model was dropped.

Lastly, the remaining three models (the two- and four-class GMMs with fixed intercepts, and the three-class GMM with fixed slopes) were compared in terms of the BIC and SSABIC values (see Table 1). The BIC and SSABIC values for the three models were 26065.64 and 25951.91 for the two-class GMM with fixed intercepts, 25889.97 and 25757.28 for the four-class GMM with fixed intercepts, and 25840.05 and 25716.84 for the three-class GMM with fixed slopes. As mentioned in the analysis section, a smaller value on the BIC and SSABIC indicates a more parsimonious model. Thus, both BIC and SSABIC values were smallest for the three-class GMM with fixed slopes. The entropy value for this model was lower than the other two, but still satisfactory. The only problem with this model is that Class 3 contained only three members, but the three students appeared to have a characteristic developmental trajectory that was distinct from the others (see Fig. 1). Therefore, it seemed reasonable to retain this class. Thus, the three-class GMM with fixed slopes was adopted as an optimal model.

Panel A of Fig. 1 shows the observed trajectories of text length for all the students, and Panel B shows the observed trajectories and estimated means for the three latent classes identified with the optimal model. The three graphs in Panel B suggest that the three classes have a distinct pattern of development. The estimated growth parameters (intercepts and slopes) of the three classes are listed in Table 2. The three classes have similar intercepts, and there was no significant difference between classes in terms of either the intercept ($F(2, 102) = .99$, $p = n.s.$) or text length of the first composition ($F(2, 101) = 2.42$, $p = n.s.$). What is noteworthy is the difference in their slopes: the slope of Class 1 has a upward trend (.97), that of Class 2 shows an even downward trend (−.13), and that of Class 3 has a marked increasing trend (3.64). The slopes between the three classes produced a significant difference ($F(2, 102) = 163.96$, $p < .001$), and follow-up tests (Dunnett's C) indicate significant differences at the .05 level between all classes. Thus, Class 1 is labeled the steadily increasing class, Class 2 the stagnating class, and Class 3 the markedly increasing class.

To further explore the differences among the three classes, the covariates were estimated to construct a conditional model (see Table 3 for the correlations among the four covariates). Table 4 lists the means of each distal outcome for the three classes and the results of Wald chi-square tests to test mean equality. The final text length significantly differed among all classes at the .001 level; the markedly increasing class tended to attain the highest text length (mean = 207.27) followed by the steadily increasing class (mean = 113.81), and the stagnating class tended to attain the lowest text length (mean = 72.94). The significant logit coefficients at the .01 level were detected for the final course grade. As mentioned in the data and indices section, the values for the final course grade were standardized within each cohort, so they represent a relative position within a cohort to which each student belonged. The average of the standardized values was 0, so for example, a negative value means a lower grade than the class average. The stagnating class tended to achieve a significantly lower course grade (mean = −.29) than the other two classes, but there was no difference between the steadily increasing class (mean = .23) and markedly increasing class (mean = .59). To summarize, there were three groups of learners who had a distinct developmental pattern of writing performance, and these groups tended to show diverged performance in not only writing and but also the course grade.

The second research question examines whether the two predictors are associated with group membership. The means and standardized deviations of the two predictors for the three classes are listed in Table 5. A conditional model with predictors as covariates was estimated, and the results are listed in Table 6. The estimates were calculated for each class in reference to another class, resulting in three comparisons between the three classes (i.e., Class 1 vs. Class 2, Class 2 vs. Class 3, and Class 1 vs. Class 3). The statistically significant logit coefficients were detected between all pairs in terms of the type of reflection (short or extended). The estimation of the steadily increasing class in reference to the stagnating class was 2.58 ($p < .01$), that of the markedly increasing class in reference to the stagnating class was 19.03 ($p < .001$), and that of the markedly increasing class in reference to the steadily increasing class was 16.45 ($p < .001$). This suggests that those who wrote extended reflections had a higher possibility of belonging to the steadily increasing or markedly increasing classes.

For the degree of variability, the comparison between the markedly increasing class and the stagnating class ($p < .01$) and that between the markedly increasing class and the steadily increasing class ($p < .05$) were significant, but no significance was found between the steadily increasing and stagnating classes. These results indicate that higher degree of variability would increase the probability of students' belonging to the markedly increasing class, but because the class contained only three students, the results should be interpreted with caution.

4. Discussion

The current study has shown that when 105 L2 learners iterated the same timed narrative-writing task procedure every week over one academic year in an intact classroom, three distinctive groups of developmental patterns in L2 writing were identified by using

Table 1
Comparison between Three Best Models (N = 105).

Fit statistics	GMM (fixed intercepts, 2 Classes)	GMM (fixed intercepts, 4 Classes)	GMM (fixed slopes, 3 Classes)
LL (No. of Parameters)	−12949.05 (36)	−12847.25 (42)	−12829.27 (39)
BIC	26065.64	25889.97	25840.05
SSABIC	25951.91	25757.28	25716.84
Entropy	.94	.93	.82
Adj. LMR–LRT (*p*)	489.52 (.001)	56.77 (.01)	94.37 (.001)
BLRT(*p*)	−13211.34 (.001)	−12877.67 (.00)	−12879.84 (.00)
Group size (%)			
C1	68 (64.76 %)	35 (33.33 %)	57 (54.27 %)
C2	37 (35.24 %)	24 (22.86 %)	45 (42.86 %)
C3		43 (40.95 %)	3 (2.86 %)
C4		3 (2.86 %)	

Note. LCGA = latent class growth analysis; GMM = growth mixture model; LL = log-likelihood value; No. of Parameters = number of estimated (freed) parameters; BIC = Bayesian information criteria; SSABIC = sample size adjusted BIC; Adj. LMR–LRT = adjusted Lo-Mendell-Rubin likelihood ratio test; BLRT = bootstrap likelihood ratio test.

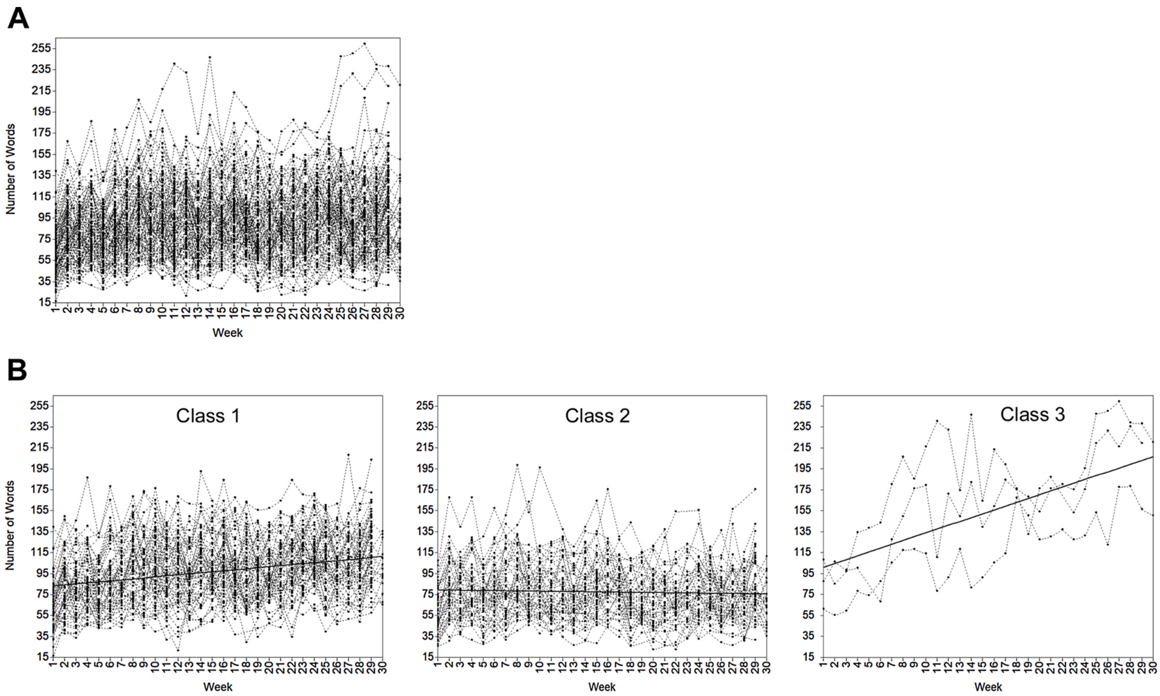

Fig. 1. Individual Text-length Trajectories for all Participants and for Members of Each Trajectory Class.
Note. Panel A: Observed individual trajectories of text length for all participants. Panel B: Observed individual trajectories of text length for Class 1 (steadily increasing class), Class 2 (stagnating class), and Class 3 (markedly increasing class) and estimated means of three classes (solid lines).

Table 2
Estimated Intercepts and Slopes for Three Latent Trajectory Classes Estimated using GMM with Fixed Slopes (N = 105).

	Class 1 (steadily increasing, n = 57)	Class 2 (stagnating, n = 45)	Class 3 (markedly increasing, n = 3)
Intercept	83.05	79.25	101.13
Slope	.97	−.13	3.64

Table 3
Intercorrelations for Four Covariates (N = 105).

	1	2	3	4
1 Final text length	–			
2 Final course grade[a]	.20*	–		
3 Type of reflection	.63**	−.00	–	
4 Degree of variability	−.30**	−.12	−.18	–

[a] Z scores within each cohort were used.
* $p < .05$.
** $p < .01$.

GMMs: the stagnating, steadily increasing, and markedly increasing latent trajectory classes. The initial text length of the three classes was similar, but the stagnating class had a negative slope of −.13, meaning their text length did not change or even degraded over the period. On the other hand, the steadily increasing class had a positive slope of .97, and the markedly increasing class had a strongly positive slope of 3.64. There were only three students in the markedly increasing class, but their trajectories were markedly distinct from those of the other students.

The three classes of students made different achievements in terms of final text length and final course grade. Final text length significantly differed between the three classes (stagnating class < steadily increasing class < markedly increasing class). This result is not surprising considering that the initial text length of the three classes was similar and their slopes characteristic. An unexpected but reasonable finding was that the final course grade was significantly higher in the steadily and markedly increasing classes than that in the stagnating class. As Baba (2020) and Nitta and Baba (2018) demonstrated, students who showed considerable progress in performance on an L2 writing task tended to take a positive attitude toward the task, and if they take a similar attitude toward other tasks

Table 4
Means of Distal Outcome Variables and Effect of Class Membership on Distal Outcome Variables (N = 105).

	Steadily increasing (C1)	Stagnating (C2)	Markedly increasing (C3)	Overall Chi-square test	Significant differences across classes
Final text length	113.81	72.94	207.27	152.69***	C1 vs. C2***, C1 vs. C3***, C2 vs. C3 ***
Final course grade[a]	.23	−.29	.59	9.36**	C1 vs. C2**, C2 vs. C3*

Note. Unstandardized logit coefficients are shown. Three-class GMM with fixed slopes was used. C = Class.
[a] Z scores within each cohort were used.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 5
Means of Predictor Variables for Three Latent Trajectory Classes (N = 105).

Latent Class	Predictors	
	Type of reflection	Degree of variability
Steadily increasing ($n = 57$)	.91 (.04)	.22 (.01)
Stagnating ($n = 45$)	.61 (.07)	.22 (.01)
Markedly increasing ($n = 3$)	1.00 (.00)	.28 (.01)

Note. Standard deviations are in parentheses.

Table 6
Effect of Predictors on Class Membership (N = 105).

Predictors	Steadily increasing (Class 1)[a]	Markedly increasing (Class 3)[a]	Markedly increasing (Class 3)[b]
Type of reflection	2.58**	19.03***	16.45***
Degree of variability	4.49	30.80**	26.31*

Note. Unstandardized logit coefficients are shown. Three-class GMM with fixed slopes was used.
[a] Stagnating (Class 2) is reference class.
[b] Steadily increasing (Class 1) is reference class.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

and activities in class or even outside class, it can be speculated that they will achieve good results in the course evaluation.

If the writing task had been investigated as a variable with the assumption of a homogeneous population, the result would have been simply that the writing task had a significantly positive effect on L2 learners. Yet, this study revealed that the reality was more complex in that the population is not necessarily homogeneous. Admitting that the focal index (text length) was simple and only captures one quantitative aspect of L2 writing, it was shown that the L2 learners evolved to form diverged populations with a divergent trajectory of development with different outcomes through iterating the writing task over one academic year. As previous studies from the CDST approach have demonstrated (e.g., Baba & Nitta, 2014; Chan et al., 2015; Lowie & Verspoor, 2019; Nitta & Baba, 2018), even learners with similar proficiency and background will develop L2 proficiency in different ways when they are scrutinized at the individual level. However, the assessment of whether or to what degree the developmental trajectories differ partly depends on at what level their performance is examined. The current study has contributed further evidence of variability in developmental trajectories even at the group level.

This suggests that development through task engagement is concerned with factors on the part of learners (e.g., Bygate, 2018; Coughlan & Duff, 1994; Lynch & Maclean, 2000). That is, iteration of a task procedure does not bring about any fixed effect but functions as *affordance* (van Lier, 2000, 2004). Nitta and Baba (2018) pointed out that "what a particular task can afford not only varies across different learners but also within the same learner at different times. Task affordances are therefore unstable and dynamically change through repeated opportunities" (p. 285), which echoes the findings of the current study. A pedagogical implication from this is that it is necessary for teachers to recognize that what a task affords to individual students in class varies and that some students may need scaffolding to learn better from the task. What functions as scaffolding may also vary depending on numerous factors such as the type of task, classroom atmosphere, and purpose in using the task, but one approach may be to let the students be aware that there are other ways to carry out the task by sharing compositions and reflective comments that other students have written. It may be also helpful to have discussions on the strengths in other students' compositions and on strategies to write a better composition.

Conditional GMMs (GGMM) with predictors and distal outcomes have shown some candidate covariates that were associated with the distinct patterns of development. The type of reflection seems to have played a role in deciding to which class each student would belong; the learners who wrote extended reflections after writing a composition tended to belong to the latent trajectory classes with a steeper positive slope. This result suggests that the way learners wrote reflections was a possible contributor to the diverged trajectories. Boud (2001) emphasized the significance of reflective practice in the learner and described reflection as "a process of turning experience into learning, that is, a way of exploring experience in order to learn new things from it" (p. 10). The extended reflection might have allowed the students in the study to explore their experiences with the task and learn more from it than the short reflection. Unfortunately, this study only focused on the effect of the two types of reflection, so future research should delve into finer analyses of leaners' reflective writing.

The effect of degrees of variability was less clear. The degree of variability of the markedly increasing class was significantly higher than those of the other two classes. These results may support Huang et al.'s (2021) and Lowie and Verspoor's (2019) findings that learners with a higher degree of variability tended to achieve more gain in their L2 writing proficiency. However, because there were only three students in the class, this result is provisional.

It should be emphasized that the current study is not experimental but observational in nature. The above results were obtained in an actual classroom context, which is, unlike a controlled laboratory context, inevitably saturated with what are traditionally considered intervening extraneous factors. However, it is of significance to conduct research in real, situated, and local contexts, as Byrne (2002) emphasizes, "[w]e absolutely need a down and dirty empiricism in which understanding is grounded in the real and constantly returns to the real" (p. 42). Such an approach, that is, finding patterns in real, complex systems, is what Byrne (2002) calls *a quantitative hermeneutic*, in which the world is understood "not through statistical modelling based on the notion of the variable, but through classification, data mining and visualization" (Byrne & Uprichard, 2012, p. xix), and this is what the current study aimed for.

The current study suggests possible foci for future research. First, because the current study was the first attempt to examine the covariates of divergent trajectory classes, a limited number of factors could be investigated. Studies at different levels and scales will be able to explore different covariates and assess their effect on developmental trajectories. Second, the effect of reflective writing was suggested, but not fully demonstrated in the current study. Only one (out of five) cohort of students wrote short reflections, so it is impossible to dismiss the possibility that the effect found was due to cohort effect. Future research should delve into not only the type of reflection but how individual learners wrote reflective comments and the subsequent influence on trajectories of writing-performance development. Lastly, the timed-writing task procedure in the current study was used for a specific purpose (students' becoming familiar with writing in English) in class in combination with other academic writing tasks, but there are two problems when used alone: it is timed and not genre-based. Considering contemporary writing pedagogy, students' performance in process-oriented and genre-based writing should also be investigated. It is hoped that findings of future studies on other aspects with other tasks will be weaved together to draw a broader and more elaborate picture in its totality to approach the complex systems of L2 writing development.

5. Conclusion

The purpose of the study was to explore whether L2 learners show divergent trajectories of L2 writing development at a group level when they iterate a narrative writing task procedure. Using a growth mixture model, three groups of learners who traced a distinct developmental trajectory were identified, and these groups made different achievements not only in the final text length but in the final course grade. These findings suggest that individuals orient and act on a task differently, resulting in different learning and divergent developmental trajectories. A conditional growth mixture model showed that learners who wrote extended reflections on their compositions after writing them tended to be classified as higher-growth groups than those who wrote short reflections. Although more elaborate analyses of how the learners write reflective comments should be conducted, the finding implies that detailed, rather than simple and superficial, reflection on one's writing may promote learning through task iteration.

Modeling L2 development is a never-ending pursuit to approximate complex systems relating to L2 learning. It should be multi-leveled and multi-scaled and require a collaborative effort to depict the essence of ever-changing complex systems. The significance of the current study lies in the contribution of a novel piece of information to this effort. It is expected that future research will not only overcome the limitations of the current study but unravel additional aspects of L2 writing development.

Funding

This work was funded by Grants-in-Aid for Scientific Research 20K00819 from Japan Society for the Promotion of Science.

Declaration of Competing Interest

The authors report no declarations of interest.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jslw.2021.100856.

References

Asparouhov, T., & Muthén, B. (2013). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling, 21*, 329–341. https://doi.org/10.1080/10705511.2014.915181

Baba, K. (2020). Exploring dynamic developmental trajectories of writing fluency: Who benefited from the writing task? In G. G. Fogal, & M. Verspoor (Eds.), *Complex dynamic systems theory and L2 writing development* (pp. 3–25). Amsterdam: John Benjamins.

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

Boud, D. (2001). Using journal writing to enhance reflective thinking. *New Directions for Adult and Continuing Education, 90*, 9–17.

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

Bygate, M. (2018). Dynamic systems theory and the issue of predictability in task-based language: Some implications for research and practice in TBLT. In M. J. Ahmadian, & M.d. P. Garcia Mayo (Eds.), *Recent perspectives on task-based language learning and teaching* (pp. 143–166). Boston: De Gruyter.

Byrne, D. (2002). *Interpreting quantitative data*. London: Sage.

Byrne, D., & Uprichard, E. (2012). Editors' introduction: Cluster analysis. In D. Byrne, & E. Uprichard (Eds.), *Cluster analysis* (Vol. 1, pp. xvii–xxxi). Los Angeles: Sage.

Byrnes, H. (2020). Envisioning L2 writing development in CDST under a curricular optic: A proposal. In G. G. Fogal, & M. H. Verspoor (Eds.), *Complex dynamic systems theory and L2 writing development* (pp. 241–270). Amsterdam: John Benjamins.

Chan, H., Verspoor, M., & Vahtrick, L. (2015). Dynamic development in speaking versus writing in identical twins. *Language Learning, 65*, 298–325. https://doi.org/10.1111/lang.12107. doi: 210.1111/lang.12107.

Coughlan, P., & Duff, P. A. (1994). Same task, different activities: The joint roles of field, purpose and engagement within the interactional approach. In J. Lantolf, & G. Appel (Eds.), *Vygotsukian approaches to second language research* (pp. 173–194). Oxford: Oxford University Press.

Diallo, T. M. O., & Morin, A. J. S. (2014). Statistical power of latent growth curve models to detect quadratic growth. *Behavior Research Methods, 46*, 357–371. https://doi.org/10.3758/s13428-013-0395-1

Fogal, G. G. (2019). Tracking microgenetic changes in authorial voice development from a complexity theory perspective. *Applied Linguistics, 40*, 432–455. https://doi.org/10.1093/applin/amx031

Fogal, G. G. (2020a). Investigating variability in L2 development: Extending a complexity theory perspective on L2 writing studies and authorial voice. *Applied Linguistics, 41*, 575–600. https://doi.org/10.1093/applin/amz005

Fogal, G. G. (2020b). Unpacking' simplex systems': Curricular thinking for L2 writing development. In G. G. Fogal, & M. H. Verspoor (Eds.), *Complex dynamic systems theory and L2 writing development* (pp. 271–294). Amsterdam: John Benjamins.

Friginal, E., Li, M., & Weigle, S. C. (2014). Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays. *Journal of Second Language Writing, 23*, 1–16. https://doi.org/10.1016/j.jslw.2013.10.001

Han, J., & Hiver, P. (2018). Genre-based L2 writing instruction and writing-specific psychological factors: The dynamics of change. *Journal of Second Language Writing, 40*, 44–59.

Howard, M. C., & Hoffman, M. E. (2017). Variable-centered, person-centered, and person-specific approaches: Where theory meets the method. *Organizational Research Methods*, 1–31. https://doi.org/10.1177/1094428117744021

Huang, T., Steinkrauss, R., & Verspoor, M. (2021). Variability as predictor in L2 writing proficiency. *Journal of Second Language Writing, 52*, 1–14. https://doi.org/10.1016/j.jslw.2020.100787

Jackson, D. O., & Park, S. (2020). Self-regulation and personality among L2 writers: Integrating trait, state, and learner perspectives. *Journal of Second Language Writing, 49*, 1–17. https://doi.org/10.1016/j.jslw.2020.100731

Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing, 12*, 377–403. https://doi.org/10.1016/j.jslw.2003.09.001

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class Growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*, 302–317. https://doi.org/10.1111/j.1751-9004.2007.00054.x

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*, 590–619.

Larsen-Freeman, D. (2018). Task repetition or task iteration? It does make a difference. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 311–329). Amsterdam: John Benjamins.

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

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research, 4*, 221–250.

Meyer, J. P., & Morin, A. J. S. (2016). A person-centered approach to commitment research: Theory, research, and methodology. *Journal of Organizational Behavior, 37*, 584–612. https://doi.org/10.1002/job.2085

Molenaar, P. C. M. (2015). On the relation between person-oriented and subject-specific approaches. *Journal for Person-Oriented Research, 1*, 34–41.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 345–368). Thousand Oaks, CA: Sage.

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*, 882–891. https://doi.org/10.1111/j.1530-0277.2000.tb02070.x

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles: Muthén & Muthén.

Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge: Harvard University Press.

Nitta, R., & Baba, K. (2018). Understanding benefits of repetition from a complex dynamic systems perspective: The case of a writing task. In M. Bygate (Ed.), *Language learning through task repetition* (p. 309). Amsterdam: John Benjamins.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535–569. https://doi.org/10.1080/10705510701575396

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford: Oxford University Press.

Tu, Y.-K., Tilling, K., Sterne, J. A., & Gilthorpe, M. S. (2013). A critical evaluation of statistical approaches to examining the role of growth trajectories in the developmental origins of health and disease. *International Journal of Epidemiology, 42*, 1327–1339.

van Lier, L. (2000). From input to affordance: Social-interactive learning from an ecological perspective. In J. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 245–259). Oxford: Oxford University Press.

van Lier, L. (2004). *The ecology of semiotics of language learning: A sociocultural perspective*. Dordrecht: Kluwer.

Van Waes, L., & Leijten, M.l. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition, 38*, 79–95. https://doi.org/10.1016/J.COMPCOM.2015.1009.1012

Vyatkina, N., Hirschmann, H., & Golcher, F. (2015). Syntactic modification at early stages of L2 German writing development: A longitudinal learner corpus study. *Journal of Second Language Writing, 29*, 28–50. https://doi.org/10.1016/j.jslw.2015.06.006

Wang, Y., & Tao, S. (2020). The dynamic co-development of linguistic and discourse-semantic complexity in advanced L2 writing. In G. G. Fogal, & M. H. Verspoor (Eds.), *Complex dynamic systems theory and L2 writing development* (pp. 49–78). Amsterdam: John Benjamins.

Wickrama, K. A. S., Lee, T. K., O'Neal, C. W., & Lorenz, F. O. (2016). *Higher-order growth curves and mixture modeling with MPlus*. New York: Routledge.

Kyoko Baba is a professor at Kinjo Gakuin University in Nagoya, Japan. Her research interests include the learning of writing skills in L2 (focusing on the instructed context), the lexical features of L2 learners' language production, and complexity theory.

Ryo Nitta is a professor and Dean of the Center for Foreign Language Education and Research, Rikkyo University, Japan. His research interests are second language writing, task-based language teaching, and language learning motivation from the perspective of complex dynamic systems theory.