

Research Article

Marjolijn Verspoor* and Kees de Bot

Measures of variability in transitional phases in second language development

<https://doi.org/10.1515/iral-2021-0026>

Received February 3, 2021; accepted February 3, 2021;

published online February 17, 2021

Abstract: This paper investigates measures of change to help demonstrate the necessity of variability as a developmental mechanism for advancing different features of L2 learning (related here primarily to writing, but also to reading) with a particular focus on learners at different stages of development. To do so, the work draws on three studies to build a case for using variability as a meaningful marker of change. Lowie, Wander M. & Marjolijn Verspoor. 2019. Individual differences and the ergodicity problem. *Language Learning* 69. 184–206 found in a group of 22 Dutch learners of English that the Coefficient of Variation (CoV), rather than individual factors such as motivation and aptitude, showed a significant correlation with writing proficiency gains. A replication study by Huang, Ting, Rasmus Steinkrauss & Marjolijn Verspoor. 2020b. Variability as predictors for L2 writing proficiency. *Journal of Second Language Writing*, with 22 Chinese learners of English revealed that the CoV rather than motivation, aptitude or working memory was a significant predictor in writing proficiency gains. A study by Gui, Min, Xiaokan Chen & Marjolijn Verspoor. Submitted. The dynamics of reading development in English for Academic Purposes, on reading for academic purposes with 27 Chinese Chemistry majors showed that the Standard Deviation of differences (SDd) rather than proficiency in English or knowledge of Chemistry correlated with reading gains. Two further studies present tentative evidence that these changes take place especially at transitional phases while learning a new skill.

Keywords: coefficient of variation; gains in proficiency; second language development; standard deviation of differences; variability

***Corresponding author: Marjolijn Verspoor**, Department of English Language and Culture, University of Groningen, Grote Kruisstraat 2/1, 9712TS Groningen, The Netherlands, E-mail: m.h.verspoor@rug.nl. <https://orcid.org/0000-0001-8828-8405>

Kees de Bot, University of Pannonia, Veszprém, Hungary, E-mail: c.l.j.de.bot@rug.nl. <https://orcid.org/0000-0001-7442-2693>

1 Introduction

The current paper is a departure from typical CDST studies that trace individuals over time in that it explores to what extent different degrees of variability among learners can tell us something about the developmental process. The focus of this paper is to explore the efficacy of tools to measure variability over time in group studies. Reviewing several case studies, we will show that those learners who have relatively higher degrees of variability make relatively more gains in writing or reading; however, it is argued that this effect may be found only when learners are changing rapidly in a particular skill. Before reviewing the actual case studies, we will review the literature on variability to explain what its function is in the developmental process and why the CoV, as a general indicator of variability, may be an adequate measure for CDST inspired group studies. However, the CoV is also affected by the slope of gains and does not take the time dimension into account and therefore the SDD is a good alternative. Finally, in our conclusion, we will argue that variability is not a personality trait nor a cause, but rather an outcome of behavior. Still, we reason that the more variable learner is also more innovative, creative, and able to adapt new strategies when needed than their less variable peer.

2 The function of variability

While variability and variation in second language development (SLD) have been studied mainly from a sociolinguistic perspective, recently a more socio-dynamic approach has gained interest. As Ortega has argued

Style: Indent 2em
Variability is thought to be an inherent property of the [SLD] system and increased variability is interpreted as a precursor for some important change in the system. The novel perspective calls for the use of new analytical methods that are quantitative, as in the traditional perspective, but also innovatively different because they are stochastic and non-causal, that is, based on probabilistic estimations that include the possibility of random variations and fluctuations tracked empirically over time. (2011: 178)

In this contribution, we will review several studies providing evidence for the idea that differences in variation in learners may be related to differences in degrees of variability over time in the development of a second language.

Variability has been investigated successfully in L2 developmental studies since the 70s of the last century (for an overview see conclusion in Pienneman 2007) but especially since the beginning of this century, it has become an object of study in its own right (cf. Lowie et al. 2021). Larsen-Freeman (1997) has been on the forefront to point out that language is a dynamic, complex system and that its

development is a dynamic, complex process. Thelen and Smith (1996) were among the first to apply a CDST perspective to developmental psychology, arguing that development should be considered as a self-organizing process, where change is a transition from one rather stable state to another. What matters here is the transition between states, not so much the states themselves. The system usually settles on one configuration out of many possible states, but during a period of transition from one rather stable state to the other, behavioral variability is essential as “the system is free to explore new and more adaptive associations and configurations” (Thelen and Smith 1996: 145). Thus, variability—the outcome of an individual, rather erratic discovery process—is considered the harbinger of change. The learner must discover, try out, and practice each part of the process him or herself, a behavioral process usually accompanied with lots of trial and error. This process is highly functional.

In CDST studies, it is argued that the degree of variability gives us information about the process and does not have to be explained in terms of causes. The cause and effect relationship between variability and development should be considered reciprocal: variability allows for flexible and adaptive behavior and is needed for development (cf. Verspoor and van Dijk 2013). In other words, there is no new behavior if there is no variability. It is the free exploration of performance that generates variability. When a learner tries out a new task the system becomes less stable, which leads to an increase in variability. Therefore, the claim is that stability and variability are indispensable aspects of human development.

In a similar line of thinking, Siegler (2006) points out that variability is normal in development. In his summary of 20 studies investigating children’s learning in micro-genetic and longitudinal approaches, he concludes that especially early on in development, the learner discovers new approaches or strategies, and that when the learner uses them, the strategies are generally used inconsistently. Thus, at early stages of development one can expect relatively more variability. Secondly, he argues that learning reflects the addition of new strategies, with greater reliance over time on relatively advanced strategies, improved choices among strategies, and improved execution of strategies. The choices of strategies are less random and would therefore result in relatively less variability. Finally, although there is variability in the process of learning, learning tends to progress through a rather regular sequence of stages. In other words, even though learners tend to follow their own idiosyncratic developmental paths—causing variability—most will improve by using more advanced strategies more consistently, resulting in relatively less variability in each learner and concomitantly less variation among learners.

Siegler’s views on strategies, albeit conceptualized in learning from a psychological view, are in line with SLD research on the use of learning strategies (see Oxford 2017 for a congruent overview). Successful learners generally use a greater

number and a wider variety of learning strategies (McDonough 1999). In contrast, a large Internet survey suggested that advanced learners used more strategies, but not necessarily a wider range (Habòk and Magyar 2018).

To explore change in individual development, longitudinal data are required with as many data points as possible to observe the change over time, relative to the rate of change of the particular behavior under investigation (cf. de Bot 2015). For example, for very specific skills that can be mastered relatively fast, such as learning to spell a frequently occurring L2 word, 20 writing samples over the course of a week may need to be traced, whereas for general skills that develop slowly such as in general writing or reading proficiency, 12 measurements over one year may be enough. In their seminal article on methods to investigate variability, van Geert and van Dijk (2002) propose several methods such as the min-max graph to visualize changes in the degrees of variability and random sampling methods to test critical moments of change (also called phase shifts) in individual trajectories or time series. Visualization techniques such as min-max graphs are meant to get a first impression of the developmental trajectory and the overall degree of variability. They can give us an indication of where to look for meaningful changes in variability, which in turn can be tested by means of random sampling methods. For L2 developmental data, these techniques have been explained and illustrated in van Dijk et al. (2011) and used in quite a few studies (e.g., Penris and Verspoor 2017; Spoelman and Verspoor 2010; Verspoor et al. 2008), which in turn have been critically reviewed by Bulté and Housen (2020).

van Geert and van Dijk (2002) argue that even though their techniques are mainly applicable to time-series of individual data, they can also be applied to cross-sectional data as long as time (such as age in months) plays a role. They illustrate this with data by Blijd-Hoogewys (2008) on the development of Theory of Mind. The major difference between individual and group data is that the variability does not apply to fluctuations within a child but to differences between children of similar and different ages. Such an approach was used on SLD writing data by Verspoor et al. (2012), which will be discussed in more detail below.

In this paper, our goal is not to trace individuals over time and discuss the methods and techniques used in such studies, but we would like to illustrate the role of variability, especially in transitional phases. If it is true that variability is a precursor for change as the learner has to go through trial and error, then one might assume two things:

- (1) If two rather similar learners, with similar initial conditions, are both going through a transitional phase, the one who shows more variability (in terms of new modes of behavior) is more likely to change over time.

- (2) Because variability is more likely to occur in very unstable systems, learners in a rapid developmental phase may show relatively more variability than learners who have reached a more stable phase. Thus, one might expect a group of beginners to be more different from each other than learners in a more advanced group of learners, whose language systems have stabilized more because constructions at all levels may have become more entrenched.

In this paper we review several case studies that support these two assumptions. However, as this is a paper on methods to test CDST claims, we first focus on the type of measures we have used and then we discuss several studies in which this measure has supported our assumptions.

3 Measures of variability

As van Geert and van Dijk (2002) point out the best-known traditional measure of variability is the standard deviation (SD), defined as the square root of the variance, which is in turn the average of the squared deviations from the mean. However, when we compare different data sets, the SD does not work well as the SD is sensitive to the mean in a sample: a higher SD is usually related to that of a higher mean. To solve this issue, the CoV is often used. The CoV is defined as the standard deviation of a sample divided by its mean—a normalized measure of the dispersion of a probability distribution. van Geert and van Dijk (2002) argue that the CoV is not totally insightful when analyzing individual variability over time, mainly because of heteroskedasticity, which is a statistical problem that is common if one deals with growth data with very low initial values (Kmenta 1990). Such low values are unstable because small absolute fluctuations are large in proportion to the values themselves and may not reflect the individual growth pattern adequately. This problem occurs naturally in child or second language acquisition at the very initial stages, where certain types of utterances may be few and far in between. However, in the studies we review below, the CoV is used for different purposes: not to see when a particular learner showed a critical moment of change (which can be calculated with various other tools), but to see if a relatively greater degree of variability leads to relatively higher gains. The CoV then is a general measure of degree of variability in a time series. It cannot tell us when something changed, but it can tell us whether in one time series or the other relatively more variability has occurred, which can then be related to change over time in terms of gains in a particular variable among members of a group.

A problem with the CoV is that it is affected by the slope and does not really take the time dimension into account, and in the last paper reviewed a different

measure was used: the standard deviation of differences. In line with Pettitt (1980) and Taylor (2000), it was operationalized as the standard deviation of differences between the raw scores and its own average on the basis of the preceding difference (SDd). Different from the standard deviation, the SDd takes the time order of the raw scores into account. Thus, differences were calculated by: $d_1 = x_1 - \bar{x}$; $d_i = d_{i-1} + (x_i - \bar{x})$. The SDd yields one indicator of the variability of a set of scores.

4 Variability as predictor of gains in transitional phases

There are three independent studies so far that have shown that degree of variability can be related to developmental differences among groups of students. The first study is by Lowie and Verspoor (2019). The original purpose of the paper was to see if an individual factor—motivation or aptitude, controlled for beginning English proficiency and out-of-class exposure—could predict the gains the learners made in L2 English in the course of one academic year. They traced 22 young Dutch learners of English who wrote between 20 and 23 short texts each. These texts were rated for proficiency level holistically as in Verspoor et al. (2012). Texts were anonymized and fully randomized for student and time of writing, and trained raters gave scores on a 5-point scale, with 1 representing the relatively weakest and 5 the relatively strongest piece of writing of the set. To control for possible strong topic effects, they checked for potential outliers, but none were found. Over time, the average ratings increased gradually from around 2.1 to 2.9. To measure proficiency gains, they calculated the difference between the average score of the first two texts and the average score of the last two texts. In a regression analysis, they tested which individual factor might be related to proficiency gains and they found that neither motivation nor aptitude could predict gains in proficiency, even when controlling for out-of-school exposure and starting level of English. However, quite by accident, in retrospect, they found that a triad of high gainers seemed to show more variability over time than a triad of low gainers.

Figures 1 and 2 show two triads of learners who were as similar as possible in terms of motivation, aptitude, starting proficiency and exposure. The first triad had the highest scores in the class on these individual factors and the second group had intermediate scores. Despite their similarities in individual difference factors, their developmental paths were quite different. The only thing the authors noticed was the different degrees of variability. The individuals in the higher group seemed to have more ups and downs in holistic scores on their texts than the intermediate triad.

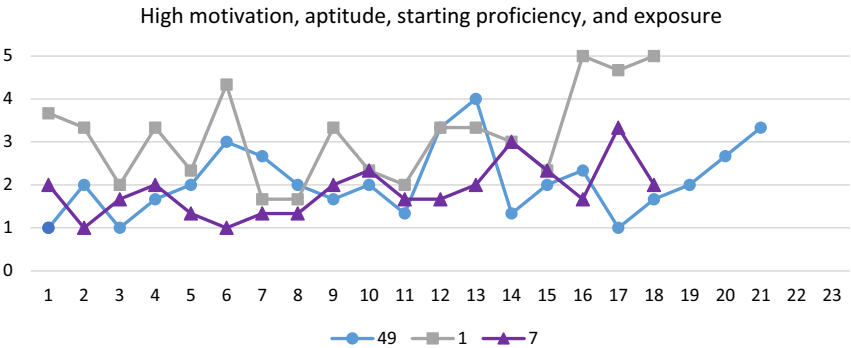


Figure 1: Data of highly similar students in Triad 1 (adapted from Lowie and Verspoor 2019).

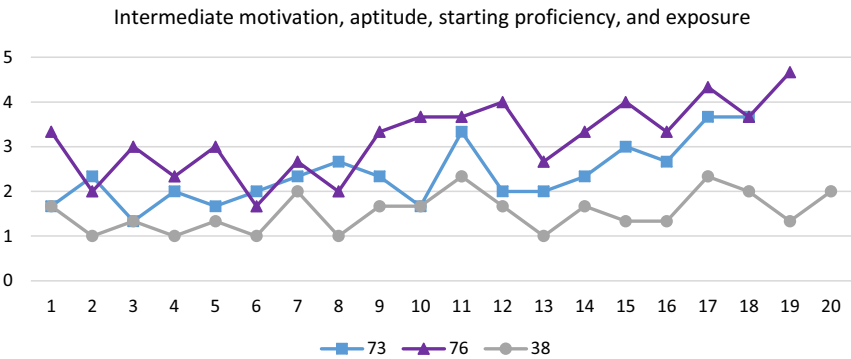


Figure 2: Data of highly similar students in Triad 2 (adapted from Lowie and Verspoor 2019).

To test this assumption, they calculated the CoV of the ratings and, in spite of the small group sizes in these triads ($n = 3$), the ratings for Group 1 were significantly more variable ($CV = 0.36$, $SD = 0.03$) than for Group 2 ($CV = 0.27$, $SD = 0.1$) ($t[4] = 3.5$, $p < 0.05$, Cohen's $d = 2.8$). Then for all participants in the experiment, correlations were calculated between the CoV and the global proficiency gains. The correlation turned out to be moderately strong positive and was significant ($r_{xy} = 0.53$, $p < 0.05$). A higher degree of variability coincided with higher overall proficiency gains.

As the findings by Lowie and Verspoor (2019) were surprising and unique, a replication study was conducted by Huang et al. (2020b). The English writing proficiency of 22 L1 Chinese adults at the university level was measured in the course of one academic year. The learners wrote 12 texts, which were scored holistically as in Lowie and Verspoor (2019), and they conducted a correlation analysis between CoV and gain scores; in addition, they included the CoV in two regression analyses,

one with final L2 proficiency as dependent variable and one with L2 proficiency gains as dependent variable. In both regression analyses, the predictors were motivation, aptitude, working memory, starting L2 writing proficiency and degree of variability operationalized as CoV. The findings were that none of the individual factors predicted the final L2 writing proficiency nor the L2 writing proficiency gains; only the CoV did.

The third study (Gui et al. submitted) is not one on a productive skill such as writing, but on a receptive skill: reading in an English for Academic Purposes (EAP) context. This study investigated the developmental trajectories of 27 Chinese students majoring in Chemistry who were enrolled in an EAP reading class. All students took a version of 1 of 12 validated and calibrated reading tests every week during one semester and each student was interviewed after each test on the use of strategies to improve on the reading tests. The SDd of the 12 texts was calculated per student to measure degree of variability over time. The group as a whole gained significantly in EAP reading, the degree of variability correlated strongly and positively with reading gains.

What was especially interesting was the complementary findings from the qualitative interview data. Table 1 summarizes in four time slots, the differences between what the top 30% students in gains and the bottom 30% students in gains mentioned related to the difficulties they faced during the test, strategies they had used since the last test to improve, and the progress that they felt they had made. The high gainers, who all had higher SDds than the low gainers, not only mentioned the use of more strategies than the low gainers, but they actually used more of them and more varied ones. Moreover, they seemed to progress more in the type of strategies they choose to use, going from the word level to the discourse level. This is very much in line with Siegler's (2006) remarks that learning reflects the addition of new strategies, with greater reliance over time on relatively advanced strategies, improved choices among strategies, and improved execution. It also is very much in line with findings in SLD research into strategies (McDonough 1999; Oxford 2017).

To summarize, we believe that these three studies show that variability plays a role in development, especially in the early phases of development of a particular skill. These three studies dealt with learners in a relatively new context. The first study dealt with young learners in their very first year at a bilingual school, the second study dealt with students in their very first year of university, and the third study dealt with students who were taking their very first academic reading course in their field of specialization, Chemistry. It is in this early phase of a new challenge that variability may be related to progress as two other studies show that such higher degrees of variability wane as learners develop. In our conclusion, we will speculate on what personality traits may be involved in higher degrees of variability.

Table 1: Interview summary (from Gui et al. submitted).

Weeks		High gainers (n = 9)	Low gainers (n = 9)
1–3	Difficulty	Technical vocabulary (24); chemistry knowledge (4)	Technical vocabulary (20)
	Strategy	Reread (10); guess word meaning (10); scan to find particular info (9); employ chemistry knowledge (9); skim before reading carefully (6); skip unfamiliar words not critical to understanding (5)	Guess word meaning (7); scan to find particular info (5); skim before reading carefully (4)
	Progress	Become familiar with EAP passages in Chemistry (13)	Become familiar with EAP passages in Chemistry (7)
4–6	Difficulty	Long and complex sentences (15); technical vocabulary (9); sentential logic (4)	Technical vocabulary (17); long and complex sentences (5)
	Strategy	Ask help from instructors (10); compare their own performance with others (9); read aloud technical words (9); collect long sentences and restudy them (8); organize technical words with tables (5); organize technical words according to the number of carbons and functional groups (4); use examples to help understanding (3); translate into Chinese (2)	Guess word meaning (5); underline keywords (3); reread (2); scan to find particular info (2)
	Progress	Know more technical words in EAP passages (15); read more fluently (12); understand the details better (12)	Understand details better (8)
7–9	Difficulty	Sentential logic (9); discursive logic (6); long and complex sentences (6); technical vocabulary (4)	Technical vocabulary (17); long and complex sentences (5)
	Strategy	Collect long sentences and restudy them (8); summarize main idea of the text (6); predict the content and check prediction (4); analyze the logic between sentences (3); adjust reading speed according to content (2)	Guess word meaning (3); translate into Chinese (3); skip unfamiliar words (2)
	Progress	Be confident when reading long and complex sentences (16); understand main idea and details better (14);	Read faster (3)

Table 1: (continued)

Weeks		High gainers (<i>n</i> = 9)	Low gainers (<i>n</i> = 9)
10–12	Difficulty	Sentential logic (8); discursive logic (8)	Technical vocabulary (15); sentential logic (3); discursive logic (3);
	Strategy	Identify key info and pass over examples (7); adjust reading speed according to content (4); predict the content and check prediction (4); check understanding (3)	Summarize main idea (1); underline key info (1)
	Progress	Identify key info (17); summarize theme with a graph or formula (10)	Understand main idea better (4)

5 Phases in degrees of variability

As we argue above, higher degrees of variability are expected to occur in transitional phases, when a very new skill is learned, and may be followed by more stable states. There are two group studies so far that point into this direction: a longitudinal and a cross-sectional one.

Huang et al. (2020) wanted to compare the effects of learning an English only (L2) versus learning two languages at the same time, a double major in English and Russian (L2 + L3). They traced the development in two cohorts, each over the course of one academic year. As in Huang et al. (2020b), the students wrote 12 texts over the course of their academic year, each one scored holistically on proficiency level with a Complexity, Accuracy, Fluency, Idiomaticity, and Coherence (CAFIC) rubric. An independent-samples *t*-test was used to compare the CoV of the L2 and L2 + L3 learners. The variability as measured by the CoV of the total proficiency scores and of each sub-score was analyzed for two different periods, namely the first half and the second half of the period of observation. The groups did not differ in terms of gains, but differences were found in degrees of variability in the first year, and especially in the first half of the first year in the fluency aspect. The L2 + L3 learners had a significantly higher degree of variability than the L2 learners in terms of fluency, both in the first half and the second half of the academic year, with a larger effect size for the first half ($r^2 = 0.33$) than for the second half ($r^2 = 0.13$). In other words, the L2 + L3 learners, who had just started to learn Russian intensively were less stable in English fluency development than the English majors, but this difference in degree of variability waned in the second half and did not occur among the second-year groups. The authors reasoned that learning the L3 intensively in their first half year

destabilized the L2 in some respect for a short while, even though it did not have negative effects on gains.

The second study is a cross-sectional one by Verspoor et al. (2012). As van Geert and van Dijk (2002) point out, CDST-inspired analyses can be applied to cross-sectional data as long as time plays a role. In this particular study, it was not age but proficiency levels that were taken as general stages in development, with the underlying assumption that over time most learners will go from one stage to the other. The major difference between individual and group data is that the variability does not apply to fluctuations within a child but to differences between children of similar and different ages. In this study, the writing samples of 489 Dutch learners of English in the first three years of high school were assessed for five different proficiency levels (1–5) roughly representing stages in L2 writing development. Each text was coded for 64 separate linguistic features involving sentence constructions, clause constructions, verb phrase constructions, chunks, the lexicon and accuracy measures. The aim was to show that at different stages of development, different linguistic sub-systems seemed to emerge and to infer from the findings what the changes in linguistic sub-systems across the stages might indicate about the L2 developmental process. Moreover, if each learner had to find their own way to detect and discover more advanced linguistic strategies to communicate, this would lead to relatively more variability within individuals especially at the early stages, depending on the particular skills they were focusing on.

The findings showed that for most variables, learners at the lower levels indeed showed more variation, also operationalized as the CoV, among each other than at the more advanced levels. The section on dependent clauses will be given here as an example. Each text was hand coded for the different types of dependent clauses as shown in Table 2.

Table 2: Dependent clauses.

Type of clause	Functioning as	Example from corpus
Finite adverbial	Adverbial	It was very nice and funny, <i>because we buyed all things the same!</i>
Finite Nominal	Subject, object or other nominal	I Said <i>I haven't saw them before.</i>
Finite relative	Post-modifier of noun	The most nice thing <i>I've did was mountain biking.</i>
Non-finite construction	Adverbial, nominal or post-modifier	He walked to the field and let the spider <i>go away.</i> Well, I met a boy, <i>named Name.</i> In de back of the boat were dol-phins <i>jumping in our waves.</i>

Figures 3 and 4 summarize the findings relating to the development of dependent clauses. Figure 3 shows that at level 1 very few dependent clauses were used, and at level 2 more dependent clauses were used, especially finite nominal and finite adverbial ones, but the increase was not significant. Significant jumps between levels occurred only between level 2 and 3 for finite adverbial clauses and all non-finite clauses. Between levels 3 and 4, there was a significant difference in the use of finite relative clauses. Figure 4 shows the CoVs for each of the variables, and it is clear that on the whole, there is much more variation among the beginners at level 1 and 2 than at the more advanced levels, 3–5.

These findings support the idea that learning tends to progress through a rather regular sequence of stages, in this case nominal and adverbial clauses before relative clauses and non-finite ones. The charts also show a development very much in line

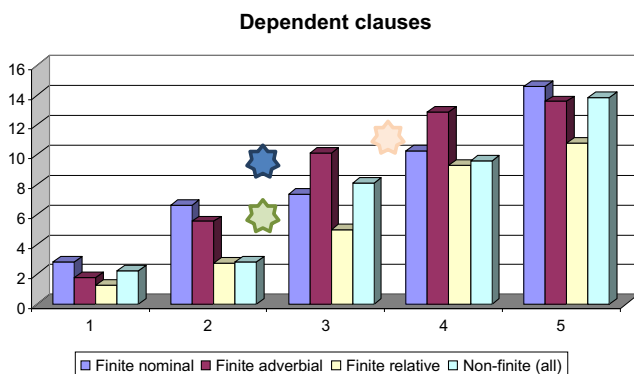


Figure 3: Distribution of types of dependent clauses at five levels of proficiency (adapted from Verspoor et al. 2012).

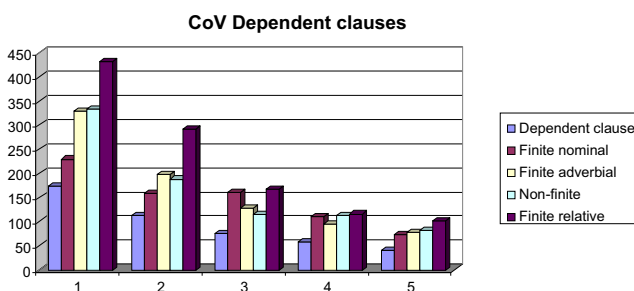


Figure 4: Coefficient of variance of types of dependent clauses (adapted from Verspoor et al. 2012).

with Siegler's findings that even though learners seem to follow their own idiosyncratic developmental paths, most will improve by using more advanced strategies more consistently, resulting in relatively less variability in learners and less variation among learners at the more advanced levels. Of course, the relatively high CoV's at the early stages are indeed probably due to heteroskedasticity in that they reflect small absolute fluctuations, which are large in proportion to the values themselves and will definitely not reflect the individual growth patterns adequately. However, it does show that the beginners are more random (only using an occasional non-finite clause or relative clause) and are more different from each other in these choices than their more advanced counterparts.

6 Conclusion

According to some of the pioneers of CDST perspectives in developmental psychology and language such as Thelen and Smith (1996) and van Geert and van Dijk (2002), variability is functional. As Thelen and Smith (1996) were perhaps the first to point out in the field of developmental psychology, increased variability is interpreted as a precursor for a change in a system: it is needed to progress. When a system is rather stable, variability occurs naturally, but at a rather low level. In contrast, the degree of variability shows an increase when a system is in a transitional phase. In other words, the occurrence of a heightened phase of variability in a learner's behavior may give us information about the fact that the learner is trying out new ways to do new things (cf. de Bot et al. 2012; Lowie 2017; Lowie and Verspoor 2015), but not always with success.

In the field of SLD many individual case studies have built on the work by van Geert and van Dijk (2002) and have explored variability patterns over time with methods such as the min-max graph to detect critical moments of change in individual trajectories or time series. The current paper did not trace individuals over time but reviewed group studies and explored the extent to which different degrees of variability among individuals could tell us something about differences in their developmental process. It was assumed that (1) learners with relatively higher degrees of variability were the ones that progressed the most, and (2) that such differences would be seen mainly at major transitional phases in development, for example when learners are just learning new skills or are in an intensive new learning phase.

To test the assumptions in the papers that were reviewed, the respective authors used a very simple and traditionally used measure, the standard deviation of a sample divided by its mean, the Coefficient of Variation. The disadvantage of the CoV, however, is that it is affected by the slope and does not take the time dimension into account, so the SDd was used in the third study. And indeed in

three independent studies among different groups of learners, degree of variability, rather than individual factors, correlated highly with gains. Even though in CDST prediction and explanation is not the exclusive goal, we may wonder what this variability means. Lowie and Verspoor (2019) suggest that “more variability may be a characteristic of a creative learning process, in which new things are tried out that may go wrong but lead to an exciting process” (202–203). Huang et al. (2020b) point out that instead of being a direct reason or actual cause of final L2 proficiency or gains, “variability should be seen as a symptom of the dynamic changes and interconnectedness of various factors in development,” and here we might speculate a little on these various factors.

From Lowie and Verspoor (2019), it was evident that although neither motivation nor aptitude turned out to be significant predictors, it was the group with the highest motivation and aptitude that showed greater degrees of variability. From Gui et al. (submitted), we learned that variability was a strong correlate of reading gains, and qualitative evidence in the interview data strongly suggests that the high gainers are much more innovative and creative, and adapt new strategies when needed. They seemed to be more ambitious and were eager to improve themselves. It is probably a mixture of individual factors—motivation, aptitude, eagerness to improve oneself and adaptability, a metacognitive capacity which can be defined as seeking new ways to do so when needed—that presents itself in their more variable behavior over time. These learners try out new ideas, which do not necessarily lead to more successful behaviors every time. Whether we can ever tease apart all these interactions of individual factors and strategies learners use is doubtful. Dörnyei (2005, 2009) has pointed out that individual factors are also a dynamic “motivation-cognition-emotion amalgam”; therefore, it would be interesting to follow up on the studies presented in the current paper with more qualitative, longitudinal data as in the Gui et al. (submitted) study to gain more insight into what drives the more variable learner.

Thus, there should be more longitudinal studies to corroborate this finding before it can be accepted as a generalizable finding. However, if replicated successfully in more studies, it will have implications for research, testing and teaching. For research, it means that general group findings about developmental sequences may never capture the actual developmental process, especially during transitional phases (cf. Larsen-Freeman 2006). As we have pointed out, group studies are very useful to find out what factors tend to play a role in development, but to learn about the actual, individual process, we need individual case studies, preferably with a mix of quantitative trajectories complemented with qualitative measures (Lowie and Verspoor 2019).

For testing proficiency levels with free response data, especially during transitional phases, the implication is that single samples should be avoided as variability

is expected, not only because of task effects, which is of course a common cause of variability (cf. Schoonen 2005), but even more so because of the fact that especially a beginner may be different from one day to the next, even on the same task. In the studies reviewed in this paper, usually the average of 2 samples was taken as the total number of samples was relatively low (12–22). However, in more dense longitudinal data, the average of 3–4 samples would be more in line with what Schoonen (2005) found was needed if samples were rated on language use. Finally, teachers should be made aware of the variable behavior of learners in transitional phases and as a result encourage learners to explore and try out new things.

References

- Blijd-Hoogewys, Els. 2008. *The development of theory-of-mind and the theory-of-mind storybooks: Lessons from a non-linear approach of developmental data in typically developing children and children with pdd-nos*. Groningen: University of Groningen.
- Bulté, Bram & Alex Housen. 2020. A critical appraisal of the CDST approach to investigating linguistic complexity in L2 writing development. In Gary G. Fogal & Marjolijn Verspoor (eds.), *Complex dynamic systems theory and L2 writing development*, 207–238. Amsterdam: John Benjamins.
- de Bot, Kees. 2015. Rates of change: Timescales in second language development. In Zoltán Dörnyei, Peter D. MacIntyre & Alastair Henry (eds.), *Motivational dynamics in language learning*, 29–37. Bristol: Multilingual Matters.
- de Bot, Kees, HuiPing Chan, Wander Lowie, Rika Plat & Marjolijn Verspoor. 2012. A dynamic perspective on language processing and development. *Dutch Journal of Applied Linguistics* 1(2). 188–218.
- Dörnyei, Zoltán. 2005. *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Zoltán. 2009. *The psychology of second language acquisition*. Oxford: Oxford University Press.
- Gui, Min, Xiaokan Chen & Marjolijn Verspoor. submitted. The dynamics of reading development in English for Academic Purposes.
- Habók, Anita & Andrea Magyar. 2018. The effect of language learning strategies on proficiency, attitudes and school achievement. *Frontiers in Psychology* 8. 2358.
- Huang, Ting, Rasmus Steinkrauss & Marjolijn Verspoor. 2020a. Learning an L2 and L3 at the same time: Help or hinder? *International Journal of Multilingualism*. <https://doi.org/10.1080/14790718.2020.1779726>.
- Huang, Ting, Rasmus Steinkrauss & Marjolijn Verspoor. 2020b. *Variability as predictors for L2 writing proficiency*. *Journal of Second Language Writing*. <https://doi.org/10.1016/j.jslw.2020.100787>.
- Kmenta, Jan. 1990. Heteroskedasticity. In John Eatwell, Murray Milgate & Peter Newman (eds.), *Time series and statistics*, 103–104. New York: Norton.
- Larsen-Freeman, Diane. 1997. Chaos/complexity science and second language acquisition. *Applied Linguistics* 18(2). 141–165.

- Larsen-Freeman, Diane. 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.
- Lowie, Wander M. 2017. Lost in state space? Methodological considerations in Complex Dynamic Theory approaches to second language development research. In Lourdes Ortega & ZhaoHong Han (eds.), *Complexity theory and language development: In celebration of Diane Larsen-Freeman*, 123–141. Amsterdam: John Benjamins.
- Lowie, Wander M., Keesde Bot & Marjolijn Verspoor. 2021. Variability as common as apple pie. *Linguistics Vanguard* 7(s2). 20200034.
- Lowie, Wander M. & Marjolijn Verspoor. 2015. Variability and variation in second language acquisition orders: A dynamic reevaluation. *Language Learning* 65(1). 63–88.
- Lowie, Wander M. & Marjolijn Verspoor. 2019. Individual differences and the ergodicity problem. *Language Learning* 69. 184–206.
- McDonough, Steven H. 1999. Learner strategies. *Language Teaching* 32. 1–18.
- Ortega, Lourdes. 2011. SLA after the social turn: Where cognitivism and its alternatives stand. In Dwight Atkinson (ed.), *Alternative approaches to second language acquisition*, 179–192. New York: Routledge.
- Oxford, Rebecca. 2017. *Teaching and researching language learning strategies: Self-regulation in context*. New York: Routledge.
- Penris, Wouter & Marjolijn Verspoor. 2017. Academic writing development: A complex, dynamic process. In Simone Pfenninger & Judith Navracscs (eds.), *Future research directions for applied linguistics*, 215–242. Bristol: Multilingual Matters.
- Pettit, Anthony. 1980. A simple cumulative sum type statistic for the change-point problem with zero-one observations. *Biometrika* 67. 79–84.
- Pienemann, Manfred. 2007. Variation and dynamic systems in SLA. *Bilingualism: Language and Cognition* 10(1). 43–45.
- Schoonen, Rob. 2005. Generalizability of writing scores: An application of structural equation modeling. *Language Testing* 22(1). 1–30.
- Siegler, Robert S. 2006. Microgenetic analyses of learning. In Deanna Kuhn & Robert Siegler (eds.), *Handbook of child psychology, volume 2: Cognition, perception, and language*, 6th edn., 464–510. Hoboken, NJ: Wiley.
- Spoelman, Marianne & Marjolijn Verspoor. 2010. Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics* 31(4). 532–553.
- Taylor, Wayne A. 2000. Change-point analysis: A powerful new tool for detecting changes. <http://www.variation.com/cpa/tech/changepoint.html> (accessed 29 November 2018).
- Thelen, Esther & Linda B. Smith. 1996. *A dynamic systems approach to the development of cognition and action*. Cambridge: MIT Press.
- van Dijk, Marijn, Marjolijn Verspoor & Wander M. Lowie. 2011. Variability and DST. In Marjolijn Verspoor, Kees de Bot, & Wander M. Lowie (eds.), *A dynamic approach to second language development: Methods and techniques*, 55–84. Amsterdam: John Benjamins.
- van Geert, Paul & Marijn van Dijk. 2002. Focus on variability: New tools to study intra- individual variability in developmental data. *Infant Behavior and Development* 25(4). 340–374.
- Verspoor, Marjolijn & Marijn van Dijk. 2013. Variability in a dynamic systems approach. In Carol A. Chapelle (ed.), *The encyclopedia of applied linguistics*, 6051–6059. Oxford: Wiley-Blackwell.

- Verspoor, Marjolijn, Monika S. Schmid & Xiaoyan Xu. 2012. A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing* 21(3). 239–263.
- Verspoor, Marjolijn, Wander M. Lowie & Marijn van Dijk. 2008. Variability in second language development from a dynamic systems perspective. *The Modern Language Journal* 92. 214–231.