

Finding the key to successful L2 learning in groups and individuals

Wander Lowie

University of Groningen, The Netherlands

w.m.lowie@rug.nl

Marijn van Dijk

University of Groningen, The Netherlands

mwgvandijk@gmail.com

Huiping Chan

University of Groningen, The Netherlands

huipingchan1981@gmail.com

Marjolijn Verspoor

University of Groningen, The Netherlands

m.h.verspoor@rug.nl

Abstract

A large body studies into individual differences in second language learning has shown that success in second language learning is strongly affected by a set of relevant learner characteristics ranging from the age of onset to motivation, aptitude, and personality. Most studies have concentrated on a limited number of learner characteristics and have argued for the relative importance of some of these factors. Clearly, some learners are more successful than others, and it is tempting to try to find the factor or combination of factors that can crack the code to success. However, isolating one or several global individual characteristics can only give a partial explanation of success in second language learning. The limitation of this approach is that it only reflects on rather general personality characteristics of learners at one point in time,

while both language development and the factors affecting it are instances of complex dynamic processes that develop over time. Factors that have been labelled as “individual differences” as well as the development of proficiency are characterized by nonlinear relationships in the time domain, due to which the rate of success cannot be simply deduced from a combination of factors. Moreover, in complex dynamic systems theory (CDST) literature it has been argued that a generalization about the interaction of variables across individuals is not warranted when we acknowledge that language development is essentially an individual process (Molenaar, 2015). In this paper, the viability of these generalizations is investigated by exploring the L2 development over time for two identical twins in Taiwan who can be expected to be highly similar in all respects, from their environment to their level of English proficiency, to their exposure to English, and to their individual differences. In spite of the striking similarities between these learners, the development of their L2 English over time was very different. Developmental patterns for spoken and written language even showed opposite tendencies. These observations underline the individual nature of the process of second language development.

Keywords: individual differences; second language development; complex dynamic systems; variability process study

1. Factors to predict L2 success in group studies

If there is one issue that the majority of researchers in second language acquisition agree on, it is the observation that individual differences (IDs) between learners are statistically associated with the success in second language learning. Differences between individuals like motivation, aptitude, and age have traditionally been treated as influential factors affecting success in second language learning. Within the long standing tradition of ID research in psychology, many studies have focused on understanding the cause of the differences between individuals in relation to learning achievement. The attention in the literature to IDs in second language development, most notably to aptitude and motivation, is still increasing. The focus on the effect of motivation alone has shown a surge in research output of the past ten years, from 33 to 138 publications, more than half of which appeared in peer review top journals in the field (Boo, Dörnyei, & Ryan, 2015). With ever more sophisticated statistical analyses, studies have attempted to identify the IDs that most accurately predict the success in learning. For instance, Gardner, Trembley, and Masgoret (1997) used structural equation modelling to identify the relative importance of a large number of IDs and explored the causal relationship between them. Using a causal modelling

approach in which they simultaneously evaluate the relationships among a large number of IDs, they show that motivation most strongly predicts achievement in the L2 (.48), followed by aptitude (.47), while confidence is most strongly loaded by achievement (.60). These and other studies focusing on the relative importance of IDs seem to agree that aptitude (the “talent for language learning”) is one of the most promising factors (prediction of success is .50), followed by motivation (.40). Another influential factor turns out to be the age of onset (.50). The relatively high correlations between success in the L2 (either based on grades or on self-assessment) and motivation is reliable and consistent, as was shown by Masgoret and Garner (2003), who carried out a meta-study of 75 independent samples. Multiple regression analyses show that the combination of aptitude and motivation, which show hardly any overlap between themselves, leads to even better prediction of success (.60). The statistical analyses have improved from simple correlations to more advanced types of analysis. For instance, using hierarchical regression analyses to determine the effect of musical ability on L2 proficiency, Slevc and Miyake (2006) find that musical ability contributes to receptive L2 phonology (.37), while age of arrival is the most important factor to predict lexical knowledge (-.42). A fully up-to-date approach to investigate the relative importance of IDs is the use of mixed effect modelling techniques in which IDs are successfully “neutralized” by including the individual as a random factor in the analysis (Kozaki & Ross, 2011; Tremblay, Derwing, Libben, & Westbury, 2011; see also Cunnings, 2012, and Linck & Cunnings, 2015).

In spite of all these promising developments, however, there have also been critical views on the relevance of IDs. For instance, Dörnyei (2009) refers to IDs as a “myth” and argues that they do not exist as identifiable factors that can contribute to success in second language learning. He disputes the major assumption that learner internal variables are independent of the environment. He argues that IDs are not distinctly definable, not stable, and not monolithic; in addition, they are strongly dependent on time and context. Dörnyei (2010) also finds that the distinction between motivation and aptitude is untenable, as illustrated by the concept of “flow,” a balanced mixture of motivation and aptitude, which demonstrates that the distinction between the two is artificial. Arguments about the non-monolithic nature of IDs have been worked out in more detail in Dörnyei and Ryan (2015), who convincingly show that the classical approach to IDs may be intuitively appealing but does not provide a realistic representation of how second language development varies as a function of time and context.

Several studies have shown that IDs are far from stable over time. Jiang and Dewaele (2015) investigated several aspects of motivation at three moments in time. Their analyses revealed a complex picture of the ideal and ought-to L2 selves, which changed over time and were affected by various motivational

variables. Significant changes occurred in the ideal-L2 self and the ought-to L2 self and their relationship with other motivational factors over the year. “The nonlinear changes in *Ideal/Ought-to L2 self*,” they show, “were consistent with the basic dynamic features of self-concept” (Jiang & Dewaele, 2015, p. 349). This study clearly showed that several ID variables interact and change over time (see Figure 1). The variable nature of IDs over time was also found in a study by Waninge, Dörnyei, and de Bot (2014) on motivational dynamics during a Spanish lesson. Even at a short timescale, which was the focus of their study (5-minute steps), motivation was highly variable and showed unique patterns of variability for different individuals in their study. We can conclude from these studies that IDs change over time at different timescales. One proposal is to redefine IDs in a more dynamic framework, as is done by Dörnyei (2009, 2010). From a complex dynamic systems theory (CDST) perspective, Dörnyei argues, higher-order ID variables can be seen as attractors that act as stabilizing forces in the developmental process. He considers ID variables in the framework of cognition, motivation and affect, and introduces factors like “possible selves” to represent individually motivated change over time. However, he also argues that there can never be a direct causal effect between these attractor states and L2 learning.

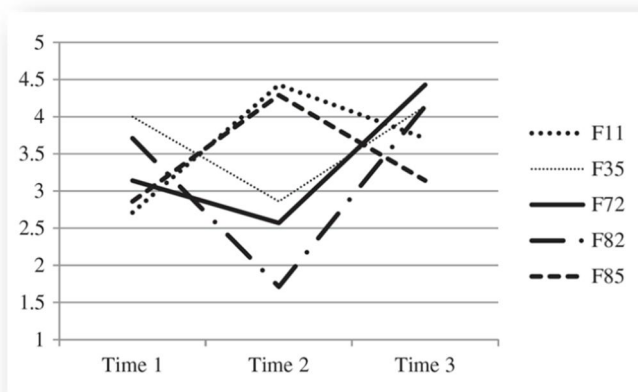


Figure 1 Variability in ought-to-self scores (1-5) for five individual participants (F11-F85) over a period of 12 months divided over three measurements (from Jiang & Dewaele, 2015)

2. Group studies versus individual case studies

In addition to the fact that IDs are not stable and delineable and may change as a function of time, there is another more serious statistical limitation to many current ID approaches. Most, if not all IDs studies have focused on inter-individual

variation and use Gaussian statistics to make conclusions about IDs based on group measures. However, such a generalization does not take into account the individual's process of development over time. This point is clearly explained by Molenaar (2015), who refers to Catell's (1952) data box. In most research, essentially two dimensions are investigated (see Figure 2). The first dimension investigates how different variables (say motivation, aptitude, and language achievement) are statically related by generalizing over observations across individuals (inter-individual variation, which we will refer to as *variation*). In the second dimension, the relationships of variables can be described in one individual case as it emerges over time (intra-individual variation, which we will refer to as *variability*). Molenaar (2015) shows that the combination of heterogeneity across subjects and heterogeneity in time violates assumptions for generalization.

Although innovations in statistical techniques are developing, most statistics currently used do not allow for generalizations across variables for different individuals in the time domain, and the analysis used is essentially a choice between either of the two dimensions. Molenaar argues that there is no relation between results obtained in statistical analyses on group data at one moment in time and an individual's development as it emerges over time, so data on the interaction of variables based on groups of individuals at one point in time cannot say anything about individual development over time and vice versa. Since most IDs have been demonstrated to be unstable and change over time, the analysis of variation will need to be complemented by analyses of variability over time.

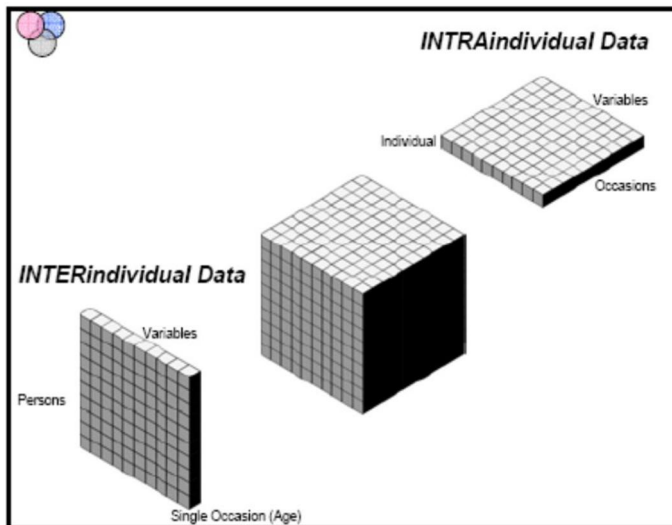


Figure 2 Catell's cube illustrating the dimensions of data analysis (Molenaar, 2015)

The analysis of variability is important if we are genuinely interested in how language changes over time in relation to IDs. In these cases, Molenaar (2015) argues in favour of subject specific data analysis for person-oriented processes. Since language development can clearly be classified as an individual person-oriented process, the combination of interacting variables and changing development must be seen as separate dimensions. One line of research is to focus on interacting variables for groups of learners, ignoring the time dimension. Virtually all studies on IDs in L2 learning have followed this line. Therefore, it is important to complement ID group studies with variability studies in which individual "differences" are excluded, but the focus is on the development over time of individual learners.

3. Degrees of variability to predict L2 success in individual learners

Thelen and Smith (1994) argue that there is not one direct cause for new behavior, but that it emerges from the confluence of different subsystems, and variability will occur in some of these subsystems because it is necessary to drive the developmental process as it allows the learner to explore and select. Because variability reflects the manifestation of the system's adaptability to the environment and signals the process of self-organization after perturbations of the system, it is a sign of development. From a more formal perspective, systems have to become "unstable" before they can change (Hosenfeld, Van der Maas, & Van den Boom, 1997). For instance, high intra-individual variability implies that qualitative developmental changes may be taking place (Lee & Karmiloff-Smith, 2002). The cause and effect relationship between variability and development is considered to be reciprocal. On the one hand, variability permits flexible and adaptive behavior and is a prerequisite to development. (Just as in evolution theory, there is no selection of new forms if there is no variation.) On the other hand, free exploration of performance generates variability. Trying out new tasks leads to instability of the system and consequently to an increase in variability. Variability is especially large during periods of rapid development because at that time the learner explores and tries out new strategies or modes of behavior that are not always successful (Thelen & Smith, 1994). Therefore, the claim is that stability and variability are indispensable aspects of human development that should be part of any analysis.

When we apply CDST insights to language development, we may assume the following: A first or second language is a complex dynamic system consisting of many subsystems such as the sound system, the grammar system, the lexical system, and so on, all of which are interrelated and may influence each other. Many internal states such as language aptitude, motivation, attitude, personality

traits, and other “individual differences” have effect on the developmental trajectory. The developmental path may further be affected by external states or events such as the general context in which a language is learned, a particular teacher, an illness, and other conditions at any given moment. All these dynamically interrelated factors may cause any part of the learner’s language system to fluctuate from one moment to the next. These fluctuations are normal for any (sub)system that has stabilized to any extent. However, strong fluctuations may indicate that a (sub)system is changing.

Learning is not linear: In both first and second language development, some subsystems may take off slowly at first, then all of a sudden jump off, and level off at the end. Other subsystems may develop in completely different ways. However, the interaction of developing subsystems will be manifested in a great deal of variability in the learner’s language. Because learners may have different starting points and learning contexts, variation among learners is also bound to exist. A great number of studies (cf. Bulté, 2013; Byrnes, 2009; Caspi, 2010; Larsen-Freeman, 2006; Murakami, 2013; Tilma, 2014; van Geert, 2008; Verspoor, Lowie, & van Dijk, 2008; Vyatkina, 2012) now have traced individual learners and shown that learners each have their own unique developmental trajectory, showing high degrees of variability and changes in variability patterns. Without explicitly mentioning it, these studies have concentrated on one individual slice of Catell’s cube, showing how variables interact in the time dimension of that individual. In these longitudinal, process-based studies with dense data, it has been found that different degrees of variability may indicate different degrees of development. For instance, high initial within-subject variability tends to be positively related to subsequent learning, and such learning reflects the addition of new strategies, greater reliance on relatively advanced strategies already being used, improved choices among strategies, and new ways to execute existing strategies (Verspoor, Lowie, & Van Dijk, 2008). For example, on a number of conservation and sort-recall tasks, children who used more and different strategies on the pre-test used more advanced strategies on subsequent tasks (Coyle & Bjorklund, 1997; Siegler as cited in Siegler, 2006).

These studies have concentrated on single learners; no study so far has compared two learners to explore to what extent the degree of variability in the development over time may be related to interacting variables. According to Catell’s separation of dimensions as explained by Molenaar (2015), interacting variables in the time dimension are not likely to be identical for different individuals. If the possibility of similarity between learners in the time dimension are investigated, it will have to be done with very similar learners to minimize the myriad of factors that may affect the degree of variability, such as differences in initial conditions, differences in personality and other “individual differences,”

and differences in external factors such as the kind and amount of exposure. Therefore, we focus on the developmental pattern of identical twins that have grown up in an identical environment and have been exposed to identical L2 input.

4. A case study of identical twins

As in Chan, Verspoor, and Vahtrick (2015), we compare identical twins, who were very similar in many respects. They live in the same home and have attended the same school in the same class. Most traditional twin studies investigate the effect of genetic factors by comparing monozygotic (MZ, or identical) twin pairs with dizygotic (DZ, or fraternal) twins (Segal, 2010; Stromswold, 2006). The current study does not focus on the genetic effect and does not compare the two types of twins but examines only one pair of MZ twins. The majority of twin studies focusing on linguistics have found identical twins to perform more similarly than fraternal twins, which validates the identical nature of their genetic makeup in the current study (Stromswold, 2006). In stating that the participants are identical twins, we are not invoking the much-maligned equal environments assumption (Plomin, Defries, McClearn, & McGuffin, 2008), which argues that MZ and DZ twins share equal environments, so any significantly closer developmental patterns found in MZ twins must be due to genetics. Instead, we merely assume that twins who share 100% of their genes and who have been raised in an identical environment are more likely than any other pair of learners to exhibit similar developmental patterns (Hayiou-Thomas, 2008). Chan et al. (2015) investigated their developmental stages over several syntactic complexity measures in both their speaking and their writing to see whether the sequences of observed developments in writing and speaking occur simultaneously or in a different order, and whether the twins develop in a similar manner. The finding was that abilities tapped by different measures developed in the spoken language before the written language and that the stages in the twins were not the same.

In the current paper, we will re-examine the data to answer our main research questions:

1. Can the degrees of variability in individuals be associated with L2 success in individual learners?
2. Can similar interactions of variables be detected in the developmental patterns of two highly similar individuals?

To be able to answer these questions, we will first investigate the development of two variables in lexical and syntactical development in both written and spoken free production tasks. The specific sub-questions pertain to each of the four variables:

1. Is there a difference between the average scores of the twins?
2. Is there a significant increase or decrease in each individual time series of scores?
3. Is there a difference in the amount of variability between the twins for all variables?
4. Is there a changing slope in the range of variability in the time series of each of the learners?

5. Method

5.1. Participants

Gloria and Grace (not their real names) are two female identical twins, aged 15 at the time of the study. For ten years, they attended school in Taiwan in the same English class with the same English teacher, where English classes were taught in Chinese with a focus on grammar. In other words, until the current study began, they had mainly received only written input in English. At the beginning of the study, they had a very similar English proficiency level (see Table 1) as measured by the General English Proficiency Test (GEPT; Wu, 2012).

Table 3 English proficiency scores (GEPT) for the twins

	Grace	Gloria
Listening (120)	112	112
Speaking (100)	80	80
Reading (120)	108	105
Writing (100)	88	82

As shown by an informal personality test, the big five test,¹ carried out at the onset of the experiment, the two girls also had similar personalities; they were rather strongly sociable, friendly, and talkative. The individual scores for the participants are represented in Table 2.

Table 4 Big five personality test for the twins (percentiles)

Trait	Description	Gloria	Grace
Openness to Experience/Intellect	High scorers tend to be original, creative, curious, complex; low scorers tend to be conventional, down to earth, characterized by narrow interests, uncreative.	7	10
Conscientiousness	High scorers tend to be reliable, well-organized, self-disciplined, careful; low scorers tend to be disorganized, undependable, negligent.	21	8
Extraversion	High scorers tend to be sociable, friendly, fun loving, talkative; low scorers tend to be introverted, reserved, inhibited, quiet.	79	79

¹ The big five personality test available at <http://www.outofservice.com/bigfive/>

Agreeableness	High scorers tend to be good-natured, sympathetic, forgiving, courteous; low scorers tend to be critical, rude, harsh, callous.	22	10
Neuroticism	High scorers tend to be nervous, high-strung, insecure, worrying; low scorers tend to be calm, relaxed, secure, hardy.	32	27

5.2. Materials

During the time of the data collection, the participants produced oral and written texts approximately three times a week, which was usually on Friday, Saturday, and Sunday. For each participant, 100 oral texts and 100 written texts were gathered. The topics, selected from the list of standard TOEFL tests by one of the researchers, were of the same genre. All the topics were presented to the two participants at the beginning of the study. Examples of the topics for writing and speaking are given below.

Example of a speaking topic:

“Which of the following statements do you agree with? Some believe that TV programs have a positive influence on modern society. Others, however, think that the influence of TV programs is negative. What TV programs have a positive influence? Why? What TV programs have a negative influence? Why?”

Example of a writing topic:

“Do you agree or disagree with the following statement? With the help of technology, students nowadays can learn more information and learn it more quickly. Use specific reasons and examples to support your answer.”

In order to motivate and remind the participants to obtain extra exposure to English and to do the speaking and writing tasks, one of the researchers created a private group on Facebook for the project, which only the researcher, the participants, and the parents had access to. The researcher reminded the twins every week to record themselves and to write the texts. Recordings were sent through email, and the written texts were posted in the Facebook account. To keep the participants motivated in the study, the researcher reacted to the content of each text, but no corrective feedback on form was given for either the oral or the written texts.

All texts were prepared for automatic processing in Lu's automatic syntactic complexity analyzer (Lu, 2010). The analyzer is designed to investigate the syntactic complexity in writing in second language acquisition, and 14 indices of syntactic complexity are calculated (see p. 479).

For our study, we used length of T-units as the complexity measure. A T-unit is defined as “one main clause plus any subordinate clause or non-clausal structure that is attached to or embedded in it” (Hunt, 1970, p. 4). A dependent clause is defined as a finite adjective, adverbial, or nominal clause, while non-finite verb phrases are excluded from the definition of clauses (e.g., Bardovi-Harlig & Bofman, 1989).

All oral texts (each about 200 words in length) were first transcribed by the researcher. To avoid redundancy in the oral production, filled pauses, dysfluencies (e.g., repetitions, restarts, and repairs), and utterances that did not involve linguistic meaning or form (e.g., laughter) were excluded. Then both the oral and written data were pre-processed for the analyzer, mainly to enable correct calculations, for instance by correcting punctuation. All other errors were left unchanged to keep the data as original as possible. After pre-processing, the text files were submitted one by one to the automatic processing tool to obtain the value of the syntactic measure for observation (mean length of T-unit = MLT).

For the lexical diversity in this study we used VocD (Malvern, Richards, Chipere, & Purán, 2004, p. 47). VocD is an adjusted metric for the type/token ratio (TTR), which is standardized for text length. In view of the differences in text length in the data, some of which were relatively short, VocD was used as a reliable measure of lexical diversity. VocD was measured as described in the following equation, illustrating standardization for text length:

$$TTR = \frac{D}{N} \left[\left(1 + 2 \frac{D}{N} \right)^{1/2} - 1 \right]$$

VocD is the single parameter of a mathematical function that models the falling TTR curve. The higher the D, the greater the diversity of a text, independent of text length. A computer program called VocD in CLAN (MacWhinney, 2000) provides a standardized procedure for measuring D (see Malvern et al., 2004).

5.3. Procedure

For this longitudinal study 100 written and 100 spoken language samples were collected during a period of eight months. For a different study that used these same data (Chan et al., 2014), the effect of input on vocabulary knowledge was investigated. For this purpose, the data contained manipulations of the input condition in three stages. A stage of relatively low input was followed by a stage of high input, followed by a stage of low input. According to the self-reports in the diaries of the participants, they obtained about 2 to 5 hours per week of extra input until Data point 20; 5 to 15 hours per week until Data point 56; and again 2-5 hours per week until the last data point. Although the manipulation is not relevant for the study reported here, it does illustrate that the two participants were exposed to virtually identical input during the period of recording the data.

5.4. Data analysis

First, we averaged the scores of each data series (MLT/written, MLT/spoken, VocD/written, VocD/spoken) to see if there was a difference between the girls

across the entire trajectory. Secondly, we tested for each girl whether there was a significant increase (or decrease) in the score over time. Then we looked at the degree of variability in the data. We aimed to discover whether there is a difference in the global variability (see below) between the two girls across the entire trajectory. Finally, we tested whether patterns of variability changed over time. More explicitly, we were interested to see whether there was a significantly greater degree of variability early on than towards the end or vice versa, and whether there was a global trend in the amount of variability across time.

In order to test the significance of the observed differences between the girls and increases or decreases within each time series, Monte Carlo permutation analyses were performed. This is a statistical testing procedure that estimates probabilities by randomly drawing samples from a dataset based on the null hypothesis, and comparing the empirically found values with a random resampling procedure. If the probability of finding the observed value in the output of the resampling procedure is very low (in this case below 5%), the result is considered to differ significantly from the null hypothesis model. (For more information on the use of permutation tests, see Todman & Dugard, 2001.)

In the current data, the Monte Carlo analysis was used to (a) test whether there was a difference between Grace and Gloria (for the mean level of MLT/spoken, MLT/written, VocD/spoken and VocD/written), (b) to test whether there was a significantly increasing or decreasing slope in each individual time series of scores, (c) to test whether there is a difference in the amount of variability between Gloria and Grace for all variables, and (d) to test whether there was a significantly increasing or decreasing slope in the variability (range) of each time series. All analyses were performed in Excel in combination with Pop-tools (Hood, 2004).

For the first Monte Carlo test, our testing criteria were the differences in the mean scores of MLT/spoken, MLT/written, VocD/spoken and VocD/written between Gloria and Grace (Gloria's mean minus Grace's mean). We reshuffled the data of the two participants across each other (5.000 times) to create resampled time series. This simulates results for the null hypothesis that there is no difference between the girls. From these simulated time series, we computed the difference between the two participants again and compared these to the empirically found differences.

For the second Monte Carlo test, the procedure was highly similar to the first, but in this case we took the global variability of each time series as testing criteria. This global variability was determined as the average of a moving range across five data points. This means that we took a moving window of five consecutive data points and calculated the local range (the maximal value in the window minus the minimal value in the window). The average of this moving

range was compared to the average of the moving range of simulated time series, based on the null hypothesis that there are no differences between the girls (see above). For both tests, we considered the difference to be significant when the probability that the reshuffled data produces the same (or larger) difference between Gloria and Grace as the observed difference is less than 5%.

The third and fourth Monte Carlo tests are based on the trend of each individual data series. The testing criteria were the linear slopes of each of these series. For the third test, we computed the slope of each data series and compared these to the slope of simulated individual time series. These are based on 5,000 reshuffles of each data series across time. This simulates the null hypothesis that the data points are independent on time. The fourth test follows the same procedure, but here the slope is based on the values of the moving ranges (with a moving window of five data points) that were computed the estimate the local variability. This slope shows whether this “local” variability is in- or decreasing over time. For both tests, we considered the result to be significant when the probability that the reshuffled data produced a slope similar or larger than the observed slope is less than 5%.

In order to analyze the relation in the performance between both girls and between the individual linguistic variables, we also performed Pearson correlation analyses. These are based on the observed values of each time series. Because of the number of tests we performed, we used a rather strict alpha of $p < .01$.

6. Results

6.1. MLT/written

When visually inspecting the trajectories of MLT/written, it stands out that Grace seems to be much more proficient than Gloria (see Figure 3). Both girls start out at a reasonably proficient level, at the beginning of the measurement period, and only Grace seems to increase during the measurements. It also shows a large degree of intra-individual variability, with several peaks, especially for Grace.

The Monte Carlo analysis confirmed the difference between the girls in global MLT/written. The average of Gloria is 9.967, the average of Grace is 12.866, and this difference is significant ($p < .001$). This means that Grace is generally more proficient than her sister. The results further showed that both slopes are positive and significant (Gloria: slope = 0.012, $p = .002$; Grace: slope = 0.032, $p = .001$), which means that both girls show an increase in MLT/written over time, though Grace’s slope is steeper. With regard to intra-individual variability, the local range of Grace is larger (Gloria has an average of 0.004, Grace 0.031). This difference tested to be significant ($p < .0001$), indicating that Grace

has more variability overall. We also tested whether this range increases or decreases over time (indicating a global change in amount of variability). The results show that both slopes are positive (0.012 for Gloria and 0.032 for Grace), but the increase was only significant for Grace ($p = .122$ for Gloria, $p = .009$ for Grace).

Combined, the results show that the trajectories of the girls are rather dissimilar: Grace is more proficient, has a steeper increase, and has more variability than Gloria. Her variability is also increasing over time, which is not the case for Gloria.

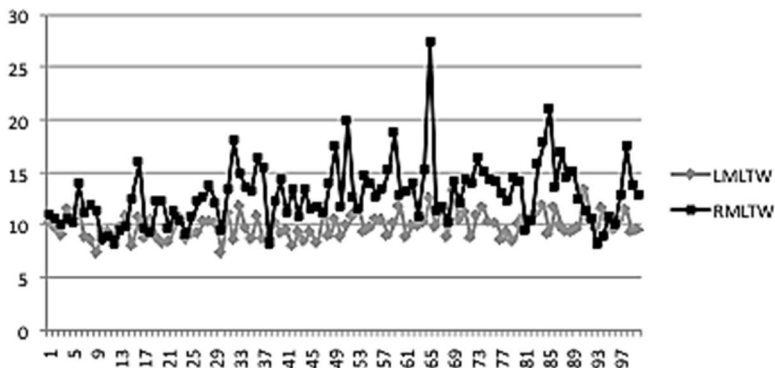


Figure 3 Written MLT for both Gloria (grey) and Grace (black)

6.2. MLT/spoken

Visual inspection the data of MLT/spoken suggest that the trajectories of the girls largely overlap (see Figure 4). Again, we observe relatively high levels of proficiency at the start of the observations and much intra-individual variability from measurement to measurement. However, it seems that the variability is more concentrated in the first half of the measurement period and decreases over time.

The results of the Monte Carlo analyses show a small difference in spoken MLT (Gloria is 13.148 and for Grace 14.204), which almost reaches significance ($p = .011$). Furthermore, the slope of Grace was significantly negative (-0.031 , $p = .006$), and nonsignificant for Gloria (0.018 ; $p = .436$). This means that Gloria's performance is relatively stable over time and that there is a slight but significant decrease for Grace. With regard to the amount of variability, the analyses show that Grace has generally more variability over the entire trajectory (Gloria has an average local range of 6.610 and Grace of 7.972, $p = .005$) and that the amount of intra-individual variability decreases over time for both. The slopes of the local ranges are negative and significant for both girls (-0.0474 for Gloria and -0.079 for Grace; $p < .001$ in both cases).

Combined, this shows that there is no general increase in proficiency in spoken syntactical development, but instead that both girls seem to stabilize. Grace's performance is somewhat more variable from moment to moment.

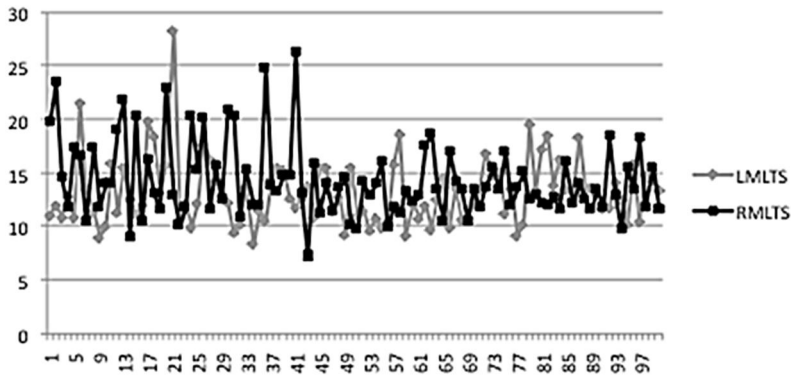


Figure 4 Spoken MLT for both Gloria (grey) and Grace (black)

6.3. VocD/written

When visually inspecting the trajectories of written VocD, the values for Gloria seem to be generally somewhat higher (see Figure 5). Again, no clear increase over time can be detected and Grace even seems to decrease over time. The amount of variability is also large again. Notably, Grace's variability seems to drop after Measurement 57.

The Monte Carlo analysis confirmed that Gloria has a higher general level of proficiency. The average for Gloria was 60.918 and for Grace 53.879, and this difference was significant ($p < .001$). Both slopes are negative (-0.052 for Gloria and -0.126 for Grace), but only Grace's is significant (p values are 0.898 and 0.001 respectively). This means that Grace's proficiency is decreasing over time. With regard to variability, no differences were found (the local range for Grace was 27.172 and for Gloria 27.595; $p = 0.615$). For both girls, there is a negative trend in local variability, indicating a general decrease of variability (Gloria's slope is -0.004 and Grace's is -0.132), but only Grace's is significant (p values are .548 and $< .001$ respectively). This means that only Grace is decreasing in her variability, indicating that her level is stabilizing.

Together, the results are somewhat different for each of the girls: Grace has a relatively low level and is decreasing over time. In addition, her variability is decreasing. Though Gloria generally shows the same patterns, they were much less pronounced and did not reach significance.

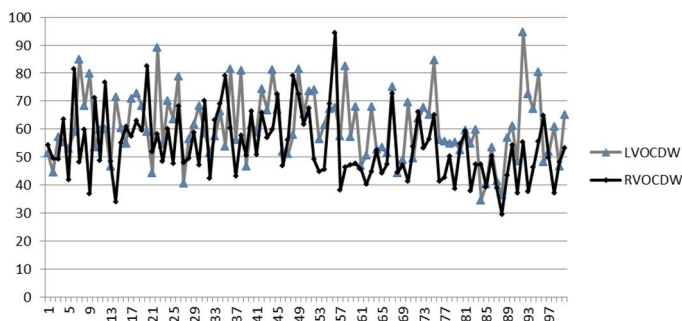


Figure 5 Written VocD for both Gloria (grey) and Grace (black).

6.4. VocD/spoken

Finally, for VocD/spoken, Gloria seems to be slightly more proficient than Grace, especially at the beginning of the trajectory (see Figure 6). Visual inspection also suggests a positive trend for Grace, and it looks like she is “catching up” with her sister. With regard to variability, this is clearly present across VocD/spoken as well, but it is hard to distinguish a clear trend.

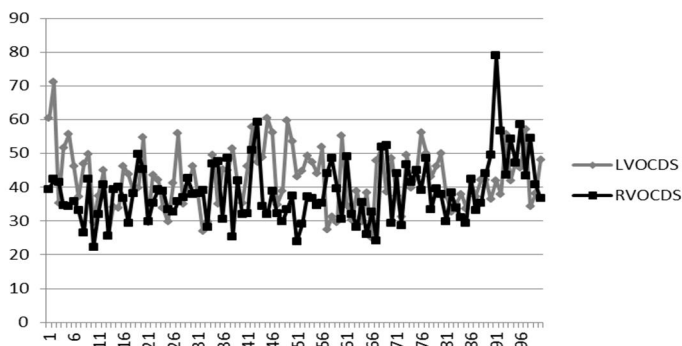


Figure 6 Spoken VocD for both Gloria (grey) and Grace (black)

The Monte Carlo analyses show that Gloria’s proficiency is indeed higher than her sister’s (the average for Gloria is 42.626 and for Grace 38.580; $p = .001$). Furthermore, only Grace has a significant positive slope (0.090, $p = .002$), and Gloria does not (-0.018 , $p = .711$). With regard to the amount of variability, there is no difference between the girls (the average local range for Gloria is 19.158 and for Grace is 17.578, $p = 0.067$). In addition, the slopes of the variability are different for each individual: Gloria’s variability is decreasing across time (-0.081 , $p < .001$) whereas Grace’s is increasing (0.078, $p = .003$).

In combination, these results show clear differences between the two girls: Grace is the one who is showing signs of development (increase in level

and increase of variability), whereas Gloria, who has an initial higher level of proficiency, only seems to stabilize over time.

6.5. Correlations

When looking at the statistical associations between the data for the two girls, the results show that there is only a significant moderate correlation between Gloria and Grace for the written VocD ($r = .297, p = .003$), and one trend towards moderate correlation for written MLT. The other correlations are not significant (see Table 3).

Table 5 Correlations between Gloria and Grace for all variables

	MLT/written	MLT/spoken	VocD/written	VocD/spoken
<i>r</i>	0.243	-0.124	0.297*	0.110
<i>p</i>	0.015	0.219	0.003	0.278

7. Discussion and conclusion

Table 4 summarizes the findings of our study. There are significant differences between the twins in both the degree of development and the degree of variability for the two variables in the two modes. The summary in the table could indicate whether one of the girls is more proficient than the other across the entire measurement period. *Significant increase or decrease in score over time* refers to a global trend in proficiency across time, that is, the average degree of variability across the entire trajectory. *Significant increase or decrease in degree of variability over time* refers to an increase or decrease in the amount of variability across time, indicating when the degree of variability would be increasing or decreasing.

Table 6 Summary of the findings of the study

	MLT				VocD			
	Written		Spoken		Written		Spoken	
	Gloria	Grace	Gloria	Grace	Gloria	Grace	Gloria	Grace
Higher average scores		X			X		X	
Significant increase or decrease in score over time	X	X		X neg		X neg		X pos
More variability in scores		X		X				
Significant increase or decrease in degree of variability over time		X pos	X neg	X neg		X neg	X neg	X pos

We observe that the patterns are dissimilar in many cases. Grace is obviously changing, but not always in the assumed direction and in different directions

in written and spoken production. She improves in written MLT but decreases in spoken MLT; she decreases in written VocD but improves in spoken VocD. She also has significantly more variability than her sister in both MLT scores. Gloria changes very little over time but increases in written MLT. She does not change in the other variables and seems to have stabilized, which is accompanied by a decreasing amount of variability in spoken MLT and spoken VocD. Also the correlation analyses showed that none of the variables strongly correlated with each other over time within each learner.

Our main research question was whether the degrees of variability in individuals might correlate with L2 success in individual learners. If we look at Table 4 we may conclude that although there is no direct one to one relation between variability and change, we may tentatively conclude that without a certain degree of variability there is little L2 change. In our data, if there is an increase or decrease, it is usually accompanied by relatively higher degrees of overall variability, and can also be seen in the direction of the slopes of variability. Variability does not guarantee success, but it does strongly seem to be a prerequisite for change to take place.

How do these case studies relate to the group studies on IDs? First of all, by controlling for as many factors as humanly possible (age of onset, general aptitude, general personality types, and so on) by investigating identical twins learning the L2 in the same environment and doing the same tasks over time, we do see remarkable differences. One of the twins is changing rather erratically in all the measures whereas the other is not. Could this have been because Grace is slightly more motivated or anxious than her sister? Even if so, it would not explain the opposite patterns for spoken and written variables.

When we link these observations to the argument in Molenaar (2015), the conclusion we can draw from our study is that Molenaar's mathematically based assumptions that observations in the time dimension need to be person specific are confirmed in the analysis of behavioral data of second language development. In this study we have explored the interacting variables in the development of two individuals, which was manifested by the amount of variability and its timing. In other words, we have investigated interacting variables at two individual slices in the time dimension of Catell's cube. In doing this, we made sure that the individual learners were maximally similar to optimize the comparability of the data. The conclusion is that in spite of the similarity of the cases achieved by minimizing IDs, very clear differences in process characteristics were found between the individual cases. This is clearly found in the data and confirmed by the correlation analysis of speaking and writing measures between the twins. Only one significant, though weak, correlation was found here.

The study into the effect of IDs on second language acquisition can focus on two dimensions. On the one hand we find evidence for the relevance of several personal characteristics that have been marked as IDs, such as motivation, aptitude, anxiety, personality, etcetera, as these variables have shown to be significantly related to achievement in second language acquisition. It has been argued that the individual variables that have been related to second language acquisition are neither monolithic nor stable, which casts some doubt on their value when they are measured at one point in time. On the other hand there is the undervalued dimension of the individual's process of development. The patterns of development emerging in individual processes are at least as revealing as the global associations coming to light in the analysis of groups of learners. In this paper we have argued that these two approaches comprise complementary perspectives as they represent different dimensions of Catell's cube (Molenaar, 2015). The relevance of the distinction between these dimensions was corroborated in our study of identical twins since even for identical twins that learn the language in identical environments, interacting variables of language development as it emerges over time are essentially different between these learners.

Many studies have attempted to crack the code to success in L2 learning by identifying IDs that are associated with the prediction of high achievement. However, the study of global differences between learners is not the only way to identify IDs as these differences can also relate to the process of learning. This process is best studied by following individual development over time. Our study of identical twins has illustrated that a focus on variability can reveal relevant and interesting differences in the individual learning process. Ideally, when advanced statistics allow us to do so, future studies should trace the interactions between variables like motivation, aptitude and achievement as they affect the process of individual development over time for groups of learners with different backgrounds and in different settings. Until that time, we should acknowledge that different dimensions of behavior need to be studied, and that the study of the process of individual development over time is at least equally revealing as group studies concentrating on interacting factors at one point in time. As Van Geert (2011) argues, "a theory of development is a theory of change, which explains how basic developmental mechanisms can generate specific developmental patterns" (van Geert, 2011, p. 276). Such a theory can provide predictions and models of developmental trajectories that single case studies can fruitfully examine.

References

- Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, 11(1), 17-34.
- Boo, Z., Dörnyei, Z., & Ryan S. (2015). L2 motivation research 2005-2014: Understanding a publication surge and a changing landscape. *System*, 55, 145-157.
- Bulté, B. (2013). The development of complexity in second language acquisition: A dynamic systems approach (Unpublished doctoral dissertation). University of Brussels, Belgium.
- Byrnes, H. (2009). Emergent L2 German writing ability in a curricular context: A longitudinal study of grammatical metaphor. *Linguistics and Education*, 20(1), 50-66.
- Catell, R. B. (1952). The three basic factor-analytic designs: Their interrelations and derivatives. *Psychological Bulletin*, 49, 499-520.
- Caspi, T. (2010). A dynamic perspective on second language development (Unpublished doctoral dissertation). University of Groningen, the Netherlands.
- Chan, H. P., Lowie, W., & De Bot, K. (2014). A case study of lexical development of writing and speaking in identical twins. In J. Gao, M. Guérin, V. Allegranzi, M. Rivière, L. Sauwala, C. Simon, & L. Xue (Eds.), *Actes des 16èmes Rencontres Jeunes Chercheurs en Sciences du Langage: Modèles et modélisation dans les sciences du langage* (pp. 54-65). Paris: HAL-SHS.
- Chan, H. P., Verspoor, M., & Vahtrick, L. (2015). Dynamic development in speaking versus writing in identical twins. *Language Learning*, 65(2), 298-325.
- Coyle, T. R., & Björklund, D. F. (1997). Age differences in, and consequences of, multiple- and variable-strategy use on a multitrial sort-recall task. *Developmental Psychology*, 33, 372-380.
- Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28, 369-382. doi:10.1177/026765831244365
- Dörnyei, Z. (2009). Individual differences: Interplay of learner characteristics and learning environment. *Language Learning*, 59(Suppl. 1), 230-248.
- Dörnyei, Z. (2010). The relationship between language aptitude and language learning motivation: Individual differences from a dynamic systems perspective. In E. Macaro (Ed.), *The continuum companion to second language acquisition* (pp. 247-267). London: Continuum.
- Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. New York: Routledge.
- Gardner, R. C., Tremblay, P. F., & Masgoret, A. M. (1997). Towards a full model of second language learning: An empirical investigation. *Modern Language Journal*, 81(3), 344-362.

- Hayiou-Thomas, M. E. (2008). Genetic and environmental influences on early speech, language and literacy development. *Journal of Communication Disorders, 41*, 397-408.
- Hood, G. M. (2004). PopTools Version 2.6.2 [Software]. Canberra: CSIRO. Retrieved from <http://www.cse.csiro.au/poptools>
- Hosenfeld, B., Van der Maas, H. L. J., & Van den Boom, D. C. (1997). Indicators of discontinuous change in the development of analogical reasoning. *Journal of Experimental Child Psychology, 64*, 367-395.
- Hunt, K. W. (1970). Recent measures in syntactic development. In M. Lester (Ed.), *Reading in applied transformational grammar* (pp. 179-192). Nueva York: Holt, Rinehart and Wiston.
- Kozaki, Y., & Ross, S. J. (2011). Contextual dynamics in foreign language learning motivation. *Language Learning, 61*, 1328-1354.
- Jiang, Y., & Dewaele, J. (2015). What lies bubbling beneath the surface? A longitudinal perspective on fluctuations of ideal and ought-to L2 self among Chinese learners of English. *IRAL: International Review of Applied Linguistics in Language Teaching, 53*(3), 331-354.
- Lee, K., & Karmiloff-Smith, A. (2002). Macro- and microdevelopmental research: Assumptions, research strategies, constraints and utilities. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 243-265). Cambridge: Cambridge University Press.
- Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning, 65*, 185-207. doi: 10.1111/lang.12117
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590-619.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*, 474-496.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Malvern, D. D., Richards, B. J., Chipere, N., & Purán, P. (2004). *Lexical diversity and language development*. New York: Palgrave Macmillan.
- Masgoret, A. M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning, 53*, 167-210.
- Murakami, A. (2013). Individual variation and the role of L1 in the L2 development of English grammatical morphemes: Insights from learner corpora (Unpublished doctoral dissertation). Cambridge University, UK.

- Molenaar, P. C. M. (2015). On the relation between person-oriented and subject-specific approaches. *Journal for Person-Oriented Research*, 1(1-2), 34-41.
- Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2008). *Behavior genetics* (5th ed.). New York: Worth.
- Segal, N. L. (2010). Twins: The finest natural experiment. *Personality and Individual Differences*, 49, 317-323.
- Siegler, R. S. (2006). Microgenetic analyses of learning. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology, volume 2: Cognition, perception, and language* (6th ed.) (pp. 464-510). Hoboken, NJ: Wiley.
- Slevc, R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, 17(8), 675-681.
- Stromswold, K. (2006). Why aren't identical twins identical? Genetic, prenatal and postnatal factors. *Cognition*, 101, 333-384.
- Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.
- Tilma, C. (2014). *The dynamics of foreign versus second language development in Finnish writing* (University of Groningen/University of Jyväskylä dissertation). Jyväskylä: University of Jyväskylä.
- Todman, J., & Dugard, P. (2001). *Single-case and small-n experimental designs: A practical guide to randomization tests*. Mahwah, NJ: Lawrence Erlbaum.
- Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61, 569-613.
- Van Geert, P. (2008). The dynamic systems approach in the study of L1 and L2 acquisition: An introduction. *Modern Language Journal*, 25, 340-375.
- Van Geert, P. (2011). The contribution of complex dynamic systems to development. *Child Development Perspectives*, 5, 4, 273-278.
- Verspoor, M., Lowie, W., & Van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *Modern Language Journal*, 92(2), 214-231.
- Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *The Modern Language Journal*, 96(4), 576-598.
- Wanninge, F., Dörnyei, Z., & De Bot, K. (2014). Motivational dynamics in language learning: Change, stability, and context. *The Modern Language Journal*, 98(3), 704-723.
- Wu, J. R. W. (2012). GEPT and English language teaching and testing in Taiwan. *Language Assessment Quarterly*, 9(1), 11-25. doi: 10.1080/15434303.2011.553251