Variability as predictor in L2 writing proficiency

Ting Huang [a,*], Rasmus Steinkrauss [a], Marjolijn Verspoor [a,b]

[a] *University of Groningen, the Netherlands*
[b] *University of Pannonia, Hungary*

A R T I C L E   I N F O

1. Introduction

The current study is a replication of Lowie and Verspoor (2019), who found that the degree of variability over time in L2 holistic writing scores correlated significantly with proficiency gains. This was a totally new and also an unexpected finding for the authors. The authors argued that this finding was in line with Thelen and Smith (1994), who argued that variability is especially large during periods of rapid development as the learner explores and tries out new strategies or modes of behaviour, which may not always be successful and thus alternate with old strategies or modes of behaviour.

In their study, Lowie and Verspoor (2019) (from now to be referred to as L&V) traced the writing development (about 22 writing samples) of a group of young Dutch learners of English for one academic year. Taking the starting L2 writing proficiency as a control variable, the regression analysis showed that none of the traditional Individual Differences (ID) factors they looked at—motivation, aptitude, language exposure—were significant predictors for the final second language (L2) writing proficiency scores. This might have been due to the fact that the group was already quite homogeneous. However, variability, operationalized as the coefficient of variation (CoV), which was calculated as the standard deviation divided by the mean of the L2 writing holistic scores, was moderately positively correlated with the overall proficiency gains, and the correlation reached a significant level. Thus, it was argued that variability is a sign of creativity or explorativity. As far as we know, the L&V study was the first to investigate whether the degree of variability was related to gains in proficiency scores, and this phenomenon must be replicated for such findings to be taken seriously.

---

* Corresponding author at: Center for Language and Cognition Groningen, Oude Kijk in 't Jatstraat 26, 9712 EK, Groningen, the Netherlands.
E-mail address: t.huang@rug.nl (T. Huang).

The current paper does so with several improvements in measures, conditions such as extra mural exposure to the L2, and types of analysis. L&V used the three major components of the L2 Motivational Self System (L2MSS, Dörnyei, 2005, 2009b; Dörnyei & Ushioda, 2009), namely Ought-to L2 self, Ideal L2 self, and Learning experience, to measure language learning motivation. The current study used a more comprehensive questionnaire designed and tested by Taguchi, Magid, and Papi (2009), which not only includes the three major component of L2MSS, but also seven other motivational factors. L&V used CITO scores (a Dutch scholastic aptitude test that most 12-year olds take before entering high school) for aptitude, which might be considered a rather broad aptitude measure. In the current study, the LLAMA test (Meara, 2005) was used to measure language aptitude (LA), which covers four aspects, i.e. vocabulary learning, sound recognition, sound-symbol correspondence, and grammar inferencing ability. Finally, L&V traced L2 English development in Dutch teenagers (12–13 years old), who were beginner learners of English, whereas the current study investigated L2 English development in 22 Chinese university students (aged 17–20) who had already been learning the L2 for years. Besides having differences in age and experience with the L2, the two groups of subjects differ in terms of L1 distance to L2 and exposure to the L2. Dutch and English are typologically rather similar languages (Schepens, van der Slik, & van Hout, 2016), and Dutch learners are exposed to a great deal of extra-mural English. In contrast, Chinese is typologically quite distant from English, and Chinese learners of English usually have very limited exposure outside of the English classroom. In the current study, we will replicate the L&V study to see if variability is a predictor in L2 writing proficiency gains.

2. Literature background

2.1. Cognitive and affective ID factors in L2 studies

ID factors have been argued to be influential in second language learning, among which the most studied ones are language aptitude, motivation, personality or language learning style, and language learning strategy (Dörnyei, 2006). They are all multi-dimensional personal characteristics which apply to everyone, and they modify and personalize a learner's trajectory of second language learning (Dörnyei, 2009a, 2010). Among these ID factors, foreign language aptitude and motivation have been argued to exert the most consistent influence on the SLA process, and they have been traditionally considered as the primary ID factors in SLA studies (Dörnyei, 2010; Dörnyei & Skehan, 2003). As suggested by Dörnyei (2010), aptitude is the most important cognitive variable, and motivation is the most powerful affective factor influencing second language learning; thus a fairly comprehensive study on ID factors should at least include these two factors.

Therefore, in the current study we investigated both the cognitive aspects and affective aspects of ID factors, i.e. language aptitude and motivation, which is also in keeping with L&V's study, but improving the instruments. As working memory has also been pointed out as one important ID and possibly a component of LA (e.g. Miyake & Friedman, 1998; Robinson, 2001, 2002, 2005; Skehan, 2012; Wen, 2016; Wen, 2019), we have included it in the current study as a separate ID factor. The three IDs investigated in this study are now reviewed in turn.

2.1.1. Language aptitude

Foreign language aptitude (LA) is an "individual's initial state of readiness and capacity for learning a foreign language, and probable degree of facility in doing so" (Carroll, 1981, p.86). Instead of being treated as monolithic, LA has traditionally been taken as a multifaceted concept consisting of four components, i.e. phonemic coding ability, associative memory, grammatical sensitivity, and inductive language learning ability (Carroll, 1965). LA has been widely studied as an ID factor influencing L2 learning, as it partly predicts language learning success (Skehan, 1991). LA was found to have a moderate positive association with L2 grammar learning (Li, 2015), and LA as measured by full-length tests strongly predicted general L2 proficiency (Li, 2016). A substantial number of studies have been carried out to investigate the impact of LA on L2 learning under different conditions. For instance, LA as measured by the LLAMA test (Meara, 2005) may significantly influence pronunciation, lexis and collocation acquisition (Granena & Long, 2012). In some other cases, LA as measured by the Modern Language Aptitude Test (MLAT, Carroll & Sapon, 1959) was shown to be the best predictor for long-term oral and written L2 proficiency when compared to the impact of first language skills and L2 affect (motivation, anxiety) (Sparks, Patton, Ganschow, & Humbach, 2009).

However, the impact of LA on L2 learning success varies depending on the situation. For example, the impact of LA varies with other internal factors such as age and external factors such as instruction method and language learning intensity. To illustrate, in Granena and Long's (2012) study, a significant correlation between LA and linguistic performance was only found for the late L2 learners. Similar results were found in Abrahamsson and Hyltenstam's (2008) study, which showed that there was a positive relationship between a high level of LA and nativelike proficiency among late L2 learners. Different components of LA were also found to have a different impact when the age of onset is involved. For example, analytic LA ability was found to be positively related to late L2 learners, while memory ability was positively related to early L2 learners (Harley & Hart, 1997). Moreover, LA was found to have a higher predictive power for language learning in explicit instruction than in implicit instruction conditions (Li, 2015). In sum, LA and the different components of LA interact with other ID factors in predicting L2 development.

2.1.2. Working memory

Working memory (WM) is a memory system which temporarily stores and manipulates the information in many complex tasks such as language comprehension, learning and reasoning (Baddeley, 1992, 2003). Similar to LA, WM is also a multi-faceted concept comprising one attentional control system—the central executive, and three subsidiary slave systems—the phonological loop, the visuospatial sketchpad, and the episodic buffer (Baddeley, 2000; Baddeley & Hitch, 1974). WM has been widely studied for its

potential impact on L2 learning in terms of different skills, such as L2 writing (Bergsleithner, 2010), L2 speaking (Mackey, Adams, & Stafford, 2010; Payne & Ross, 2005), L2 vocabulary learning (Atkins & Baddeley, 1998), and reading (Harrington & Sawyer, 1992).

Similar to LA, the impact of WM on L2 learning varies with other ID factors, such as age and L2 proficiency (Linck, Osthus, Koeth, & Bunting, 2014). Different components of WM also have been found to interact with other internal learner factors such as L2 proficiency; for instance, the role of phonological short-term memory capacity in L2 learning varied depending on the learners' proficiency level (Kormos & Sáfár, 2008). Similarly, phonological WM was found to be more influential among the lower proficiency learners (Serafini & Sanz, 2016). According to Williams (2012), an adequate WM span task should measure the ability to both retain and manage the information while potentially being distracted by other cognitive tasks. Such WM span tasks are normally complex span tasks such as the operation span task, reading span task, listening span task, and so forth (Wen, 2015, 2016). Moreover, the language used in a WM span might influence the subjects' scores depending on their language proficiency (Mitchell, Jarvis, O'Malley, & Konstantinova, 2015). Therefore, in the present study, we used the operation span task, which is both a complex task and language free, and thus particularly suitable for second language studies, as suggested by Williams (2012).

A large number of studies have been done to investigate the relationship between WM capacity as measured by the operation span task and language proficiency. Linck and Weiss (2011) studied how the WM would predict the improvement in L2 proficiency in terms of explicit knowledge as measured by vocabulary and grammar tests, within one semester. The results of the regression analysis showed that the higher WM predicted higher gains in L2 proficiency. WM as measured by the operation span task was also found to predict the accuracy concerning different aspects of language learning. For example, WM was found to correlate with novice adult learners' productive accuracy of noun-adjective agreement patterns in Russian under an explicit learning condition (Denhovska, Serratrice, & Payne, 2016). Moreover, it was found to predict reading accuracy given the effects of stress, and affect the effectiveness of different types of feedback (Rai, Loschky, Harris, Peck, & Cook, 2011).

2.2. Motivation and L2 motivational self system

Motivation accounts for why people decide to do an activity, and influences how much effort they are willing to spend, and for how long they are going to sustain it (Dörnyei, 2000). Language learning motivation is the learners' drive to learn the language, and it is a combination of desire, affection and effort toward language learning (Lalonde & Gardner, 1985). Motivation has widely been studied with respect to different L2 skills such as speaking and writing. It was found to play an important role in L2 speaking proficiency gains both for students studying abroad, who had more direct contact with the target community and culture, and students studying at home (Hernández, 2010). Students with higher motivation were also found to improve more in L2 speech comprehensibility (Saito, Dewaele, & Hanzawa, 2017). More specific types of motivation, such as L2 writing motivation, were also studied. For example, L2 writing motivation was especially predictive in terms of how the L2 writers process corrective feedback (Waller & Papi, 2017).

From an L2 Motivational Self System (L2MSS) view, motivation is multifaceted in nature. The concept of a L2MSS has drawn much attention and triggered numerous studies, but to our knowledge, there are only a limited number of studies investigating the predictive power of L2MSS on L2 proficiency or L2 proficiency gains. Among the very few studies, Lamb (2012) and Moskovsky, Assulaimani, Racheva, and Harkins (2016) investigated to what extent the tripartite L2MSS can predict L2 proficiency, and L&V examined how motivation measured in the beginning of the academic year can predict the L2 writing proficiency by the end of the academic year. In these three studies, only one study found only one sub-component of L2MSS (i.e. L2 learning experience) to be a significant predictor of L2 proficiency (Lamb, 2012), but in the other two studies, none of the three components of the L2MSS were found to predict L2 reading and writing proficiency (Moskovsky, Assulaimani, Racheva, & Harkins, 2016) or writing proficiency gains (Lowie & Verspoor, 2019). Given the seemingly disappointing results in the previous studies regarding the predictive power of L2MSS on L2 learning, the current study adopted the English learning questionnaire designed and tested by Taguchi, Magid, and Papi (2009), which covers a wider scope in motivation, hoping to encompass as many of the sub-aspects of motivation as possible.

To summarize, Dörnyei (2010) concluded that the correlations between L2 attainment indices and LA are often around 0.50 and those for motivation are around 0.30 to 0.40, which are much higher than the correlations between L2 proficiency outcomes and other ID factors such as learner styles and beliefs. In addition to the LA and motivation dyad, WM has also drawn increasing attention as an individual cognitive process underlying L2 processing and learning, and is positively associated with L2 attainment, with the estimated effect size as high as 0.255, as found in the meta-analysis conducted by Linck et al. (2014). However, as Dörnyei (2010) also pointed out, cognitive and affective factors may also affect each other over time and may change in interaction with each other and the L2 developmental process, so we should not look at any of these factors as being mono-causal and able to predict L2 development directly (Dörnyei, 2010).

2.3. Variability

Due to the interconnectedness and dynamic nature of the various ID factors, development in language is by no means linear and shows a lot of variability. Variability is an inherent property of a dynamic system, with different degrees and patterns of variability providing insight into the developmental process itself (de Bot, Lowie, & Verspoor, 2007; Lowie & Verspoor, 2015, 2019; Spoelman & Verspoor, 2010; van Geert & van Dijk, 2002). From a behavioral science view, variability is not necessarily caused by external factors; instead, it can also be produced by the self-organizing system itself, and the amount and pattern of variability could be seen as a predictor of development (de Weerth, van Geert, & Hoijtink, 1999).

Verspoor, Lowie, and Dijk (2008) demonstrated the values of regarding variability itself as a source to gain insight into the developmental process by reinterpreting the data from Cancino, Rosansky, and Schumann (1978), and analyzing the writing data of an

advanced learner of English. The analyses with moving min-max graphs (van Geert & van Dijk, 2002) showed that a higher degree of variability could indicate a transition phase. As suggested by Spoelman and Verspoor (2010), the degree of variability may change according to the stability of the system at a given moment, and should "be treated as data and be analyzed" (p.533). Their study investigated the development of accuracy and complexity in L2 learners of Finnish, and the findings suggested that, particularly for accuracy, there was a greater degree of variability in the early stage than at the more advanced stage. Thus far, degrees of variability have been studied in longitudinal individual case studies, showing that large degrees of variability indicate a transitional phase, but as far as we know, there have not been studies before L&V to show that learners who show a higher degree of variability are more likely to progress more.

In a longitudinal group study, L&V examined the predictive power of ID factors versus variability on L2 writing proficiency gains. They traced the L2 English writing development of 22 L1 Dutch teenagers (aged 12–13) with 23 texts written during one academic year. L2 writing proficiency gains were measured by the difference between the average of the first and last two holistic scores of the 23 writing texts. They found that neither motivation nor aptitude were able to predict L2 English writing proficiency gains. However, there was a strong and positive correlation between gains in proficiency and degree of variability, operationalized as the coefficient of variance (CoV). This meant that students who showed most variability gained the most. Their findings raise the question whether they can be replicated and validated, ideally in a different context and with more rigorous methodology than in the original study. The current study seeks to fill this gap by trying to answer the following questions:

1. Can the findings in L&V be replicated in a different population?

In other words, do the ID factors fail to predict final L2 writing proficiency while controlling for starting L2 writing proficiency, and does CoV correlate with L2 writing proficiency gains?

2. Do the ID factors and CoV predict the final L2 writing proficiency?
3. Do the ID factors and CoV predict the L2 writing gains?

We hypothesize that the results in L&V can be replicated, and according to the findings in L&V, we assume that none of the ID factors is a significant predictor for L2 writing proficiency (gains) in the regression analyses, but the CoV will be.

3. Methodology

This longitudinal study traced 22 Chinese students majoring in foreign language studies at university level. At this university, there are two different foreign language programs: one traditional foreign language program in which one majors in one foreign language (English) and a new bi-foreign-language program in which one majors in two foreign languages simultaneously (English and Russian in our study). This longitudinal study examined the participants' ID factors at the beginning of the academic year, and traced the participants in their English writing development during this period (9 months), resulting in 12 writing texts per participant.

3.1. Participants

The participants (21 females and 1 male) had a mean age of 18.68. Of the 22 participants, 9 were English majors from the traditional foreign language program, and 13 were English/Russian majors (henceforth E/R learners) from the bi-foreign language program. All of the participants started with an estimated B1 CEFR (Common European Framework of Reference) English proficiency regarding writing and reading skills. The E/R learners had no experience in Russian until they started the university study.

3.2. Language exposure versus language learning groups

During the academic year of observation, all of the participants had 16 h of English instruction per week, covering basic English speaking, reading, writing, and listening comprehension skills training. In addition, the E/R students had 8 h of Russian instruction per week. Unlike the participants in L&V, who had rather massive out-of-school exposure to English thanks to the school setting and the Dutch environment, the participants in the current study had very limited exposure to either English or Russian outside of the classroom. In L&V, language exposure was controlled for as a factor and was represented as an ordinal score on a 3-point scale. Given that there is very limited extramural language exposure in China, we did not control for language exposure. Instead, we controlled for learning one foreign language versus two. The main reason is that, in two related studies, we found a differential impact of these two conditions on WM, and on variability patterns over time. The E/R learners gained significantly more than the English learners in WM after one academic year (Huang, Loerts, & Steinkrauss, 2020), and showed more variability in L2 writing development than their English-only counterparts (Huang, Steinkrauss, & Verspoor, 2020). Therefore, in the current study, the language learning group was also considered as a potential factor affecting the final L2 writing proficiency and L2 writing gains.

3.3. Language aptitude measures

The LLAMA test (Meara, 2005) was used to measure participants' language aptitude, given its attested reliability and validity (e.g. Granena, 2013; Rogers, Meara, Barnett-legh, Curry, & Davie, 2017). Moreover, it is free to download online (http://www.lognostics.

co.uk/tools/llama/) and easy to administrate via a computer. The LLAMA test is loosely based on the Modern Language Aptitude Test (MLAT, Carroll & Sapon, 1959) and consists of four subtests measuring four aspects of language aptitude correspondingly (Meara, 2005).

LLAMA-B measures vocabulary learning ability by asking the task takers to learn and remember 20 objects and their associated words from an unfamiliar language in a limited time. LLAMA-D measures the ability to recognize sound patterns in spoken languages. It plays the sound of 10 artificial words, and requires the test takers to recognize them when the words are repeated together with other new artificial words. LLAMA-E measures the ability to establish sound-symbol-correspondences by displaying 24 buttons labeled with a written-out syllable in an unfamiliar notation. When pressed, each button plays the sound associated with the notation. After a familiarization phase, the test takers are asked to choose the correct notation for previously unheard syllables, requiring them to uncover the spelling pattern in this artificial language. LLAMA-F measures the grammar inferencing ability by introducing the task takers to sample sentences of an artificial language with matching pictures showing their meaning, and then requiring them to pick the correct sentence when shown novel pictures based on the grammar they inferred from the examples. All subtests have scores ranging from 0 to 100.

3.4. Working memory measures

An automatic version of the Operation Span task (OSP, Unsworth, Heitz, Schrock, & Engle, 2005) was used to measure WM capacity. The OSP task is a complex domain-general WM span task, which measures not only the ability to maintain information but also the ability to manage and manipulate the information (Wen, 2016). It differs from other complex domain-specific span tasks such as the speaking and reading span task in that it is language independent (Wen, 2016). The OSP task is therefore particularly suitable for L2 research (Williams, 2012).

The OSP task used in the present study was created in E-prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002). The test takers were shown several screens in succession. First, a simple math equation was displayed, followed by a solution to the equation that the test takers have to confirm or reject, and finally a letter was presented. This process was repeated 3–7 times after which the task takers were required to recall the letters in the order they were shown. This process concluded one set, and the test takers were shown 15 sets in succession. In this way, the OSP task assessed not only the memory span, i.e. how many letters can be recalled in the right order, but also the ability to maintain information (the letters) while being distracted by a different task (solving the equations). The absolute score of the OSP, which is the total number of perfectly recalled letters in all sets in the right order, was used in this study to represent the participants' WM score, and its possible range was from 0 to 75.

3.5. L2 learning motivation

The Chinese version of the English learning motivation questionnaire tested in Taguchi et al. (2009) study was used to measure motivation. Apart from the three core factors of the L2MSS, i.e. ideal L2 self, ought-to L2 self and L2 learning experiences, the test includes 7 more factors: intended effort to learn English, family influence, instrumentality-promotion, instrumentality-prevention, attitude toward L2 community, culture interests, and integrativeness. The 67 questionnaire items are divided into two parts, with the first part being statement-type items measured by six-point Likert scales, and the second part being question-type items measured by six-point rating scales with the anchors *not at all* and *very much*. Although the reliability of the questionnaire was tested in previous studies, for instance in Taguchi et al. (2009), the reliability of the questionnaire was further tested with a Cronbach's Alpha test for our own homogenous group of learners. See Table 1 for details.

According to Dörnyei and Csizér (2012), in a Cronbach Alpha test, we should aim for a coefficient above 0.70, and a coefficient below 0.60 is problematic. In our analysis, all factors were well above 0.60 with the exception of the factor integrativeness (0.58), which was therefore excluded from our analysis in the following steps.

To reduce the number of predictor variables for the analysis, the nine motivational factors from the 22 participants were submitted to a principal component analysis (PCA). This was done to firstly, avoid a possible over-fit of the regression model when using many motivational factors as predictors given our relatively small sample size, and secondly, to avoid possible correlations between the

Table 1
Composites of Motivational Variables with Cronbach Alpha Coefficients.

Factor name	Item number in questionnaire	Cronbach Alpha coefficients
Intended effort	3, 13, 23, 31, 37, 45	0.75
Ideal L2 self	6, 14, 29, 38, 46	0.80
Ought-to L2 self	5, 12, 19, 27, 36, 42, 49	0.87
Family influence	2, 11, 21, 30, 40	0.78
Instrumentality-promotion	4, 10, 16, 22, 28, 35, 41, 48	0.84
Instrumentality-prevention	7, 18, 25, 33, 43	0.90
English learning experience	50, 55, 60, 65	0.79
Cultural interest	53, 58, 63	0.69
Attitudes to L2 community	54, 59, 64, 67	0.77
Integrativeness	52, 57, 62	0.58

Note. Adapted from Taguchi et al. (2009).

motivational factors limiting explanatory value. The sampling adequacy for the analysis was verified by the Kaiser-Meyer-Olkin measure (KMO = 0.65), and the Bartlett's test of Sphericity turned out to be significant at the 0.001 level. While a PCA can reduce the number of variables, a rotation of the components is a necessary step in a PCA to produce components that are uncorrelated. Oblique rotation was used as the underlying original motivational factors can be assumed to be correlated, too. After oblique rotation, two components with eigenvalues over 1 explaining 71.37% of the variance were obtained from the analysis. Based on the motivational factors loading on component 1 and 2 respectively, component 1 was considered as tapping into extrinsic motivation (EM) and component 2 more into intrinsic motivation (IM). For each participant, the scores (based on the respective loads of each factor) of the two clusters served as motivation scores in the regression analysis. Details of the PCA analysis are shown in Table 2.

3.6. Variability operationalization

Variability was operationalized in the same way as in L&V, using the coefficient of variation (CoV) to represent degree of variability. The CoV is calculated as the standard deviation divided by the mean of the L2 writing scores of all texts written during the academic year. The standard deviation (sd) of the 12 texts from each participant reveals the amount of variation in the developing process, but as the sd might depend on the overall mean of the data, the sd is divided by the mean to obtain the CoV.

3.7. L2 writing proficiency measures

To measure L2 writing proficiency, written texts were collected from the participants every three weeks for a duration of one academic year, resulting in 12 texts for each participant. The tasks were the writing assignments from a comprehensive English course, in which vocabulary and grammar are introduced for reading, listening, speaking and writing skills under one main topic for each unit. The task complexity of the writing assignments increased with each study unit. For more detailed information on the writing tasks, see Table 3.

All of the writing texts were rated holistically based on the CAFIC model (i.e. complexity, accuracy, fluency, idiomaticity and coherence) proposed by Hou, Verspoor, and Loerts (2016) and validated with the current data set in two steps. First, the three authors assigned a proficiency order to nine randomly selected samples, and subsequently refined the CAFIC rubric and scores. Then the first author validated the rubric by re-rating the nine texts independently using the 1–5 scale on each aspect of CAFIC with a possible total score of 25. The correlation analysis found a significant relationship between the two ratings (r = 0.826, p < 0.01), with a large effect size.

Using the validated CAFIC rubric, the first author rated the whole writing dataset one writing task at the time. To ensure consistent rating, the rater randomly selected ten texts already rated from the previous writing task and rated them a second time. Only if the two scores were highly and significantly related did the rater start to rate the next writing task. After the whole data set had been collected and rated, the second author randomly selected 50 texts from the whole data set, and the first author rated these again. A correlation analysis was carried out on the two ratings. A significant correlation (r = 0.75, p < 0.001) between the two total scores (i.e. sum of the five sub-scores on CAFIC) with a large effect size indicates that the rater was reliable and consistent in the rating process. Therefore, the reliability of the whole rating process was ensured. In the current study, the total score of these five dimensions of CAFIC (ranging from 5 to 25) served as the writing proficiency score for each text.

4. Analysis

The scores used in the analyses were as follows: for LA, the scores of the four subtests were used; for WM, the absolute score from the OSP task was used; and for motivation, the two component scores (weighted average) were used. For variability, the CoV was used, and for L2 writing proficiency, the total score (25 max) was used. Language learning group was a nominal variable representing whether the students were English only or E/R learners. For proficiency, the starting L2 writing proficiency was operationalized as the mean score of the first two writing samples, and the final L2 writing proficiency was operationalized as the mean score of the last two writing samples. L2 writing proficiency gains were measured as the difference between starting and final L2 writing proficiency.

In order to replicate L&V as closely as possible with the current dataset, we first carried out a regression analysis with our three ID

Table 2
Principal Component Analysis on the Nine Motivation Factors.

Component	Label	Motivation factor	Component load	Extraction
1	EM	Ought-to L2 self	0.952	0.908
		Instrumentality-prevention	0.913	0.917
		Family influence	0.893	0.806
		Instrumentality-promotion	0.869	0.768
		English learning experience	0.827	0.776
		Attitude toward L2 community	0.824	0.720
2	IM	Culture interest	0.761	0.579
		Intended effort	0.699	0.512
		Ideal L2 self	0.627	0.437

Note. Extraction method: Principal Component Analysis. Rotation Method: Oblimin with Kaiser Normalization.

Table 3
Writing tasks.

Topic of the unit	Genre	Required words
Half a Day – *by* Naguib Mahfouz	Narrative	150
The Boy and the Bank Officer – *by* Philip Ross	Narrative/Descriptive	150
Message of the Land – *by* Pira Sudham	Descriptive	150
Midnight Visitor – *by* Robert Arthur	Narrative/Descriptive	150
The Nightingale and the Rose – *by* Oscar Wilde	Descriptive	150
The Green Banana – *by* Donald Bechelder	Descriptive	150
Another School Year – What for? – *by* John Ciardi	Descriptive	200–250
Say Yes – *by* Tobias Wolff	Descriptive/Argumentation	200–250
The Rite of Spring – *by* Arthur Miller	Descriptive/Agumention	200–250
The Man in the Water – *by* Roger Rosenblatt	Descriptive/Argumentation	200–250
Quick Fix Society – *by* Janet Mendell Goldstein	Descriptive/Argumentation	250–300
Wisdom of Bear Wood – *by* Michael Welzenbach	Descriptive/Argumentation	250–300

Note. Adapted from Yang et al. (2010).

factors (i.e. language aptitude, WM, and motivation), starting L2 writing proficiency, and language learning group as independent variables and the final L2 writing proficiency as dependent variable. Then, we ran a correlation analysis between degree of variability and L2 writing proficiency gains (RQ1).

However, a correlation analysis has its own limitations. It provides information on the direction and strength of a pairwise relationship between two variables; however, it does not demonstrate the predictive relation between the two variables. In contrast, a regression analysis can take more than one independent variable into account and presents an equation on to what extent each independent variable can predict the dependent variable (Bewick, Cheek, & Ball, 2003). The original L&V analysis did not test the ability of variability as a factor in a regression analysis to predict either L2 final writing proficiency or L2 writing proficiency gains compared to the other ID factors. Therefore, the current study added two regression analyses to the original design with final L2 writing proficiency as dependent variable in the first analysis (RQ2), and L2 writing proficiency gains as dependent variable in the second regression analysis (RQ3). In both regression analyses, the predictors were the three ID factors, degree of variability, starting L2 writing proficiency, and language learning group.

5. Results

Just like L&V did in their study, a preliminary investigation was performed on the overall improvement of the learners' L2 writing proficiency by comparing the average of the first two writing scores (early) and the last two writing scores (late) (see Fig. 1). The results of the paired samples *t*-test show that the participants significantly ($t(21) = 8.12$, $p < 0.001$) improved in L2 writing proficiency over time from early($M = 14.16$, $SE = 0.30$) to late ($M = 16.84$, $SE = 0.25$), with a large effect size ($r^2 = 0.76$).
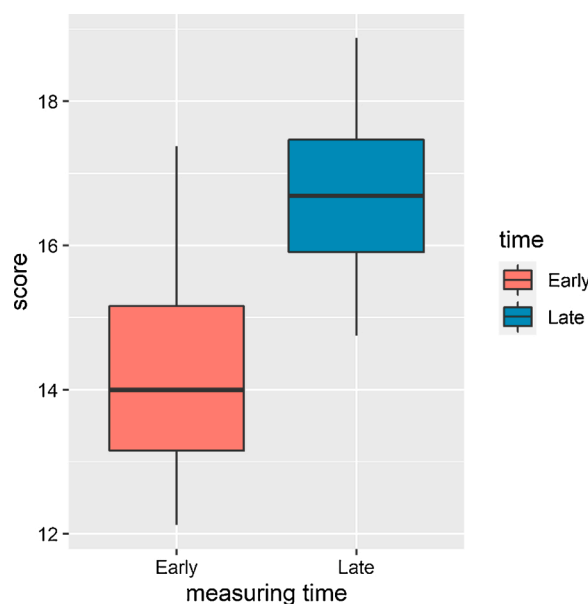

Fig. 1. General improvement in L2 writing proficiency.

5.1. Analysis 1

In our replication of the analysis in L&V, we took the traditional ID factors (aptitude, working memory and motivation), starting L2 writing proficiency, and language learning group as factors to predict final L2 writing proficiency. The multiple linear regression analysis showed that none of these factors was significant. The LLAMA-E scores (sound-symbol correspondence ability) (.45) and the starting L2 writing proficiency (.44) had the highest standardized coefficients, but did not reach a significant level (p = 0.11 and 0.12, respectively). A Pearson correlation analysis was then carried out to investigate the relationship between variability and L2 writing gains. The results indicated that CoV and gains were significantly positively correlated (r = 0.79; p < 0.001, two-tailed) with a large effect size ($r^2 = 0.64$).

5.2. Analysis 2

In our second analysis we performed a regression analysis to predict final L2 writing proficiency. In contrast to analysis 1, we included the CoV in addition to the ID factors (aptitude, working memory and motivation), starting L2 writing proficiency, and language learning group as factors. The result of the multiple linear regression analysis using the forced entry method showed that the starting L2 writing proficiency and CoV were the only two significant predictors for the final L2 writing proficiency; both predictors contributed positively. A significant regression equation was found (F(10, 11) = 2.93, p < 0.05), with an explained variance R^2 of 0.73. Table 4 shows the details.

5.3. Analysis 3

Using the same method of analysis as in analysis 2, but taking L2 writing proficiency gains rather than final L2 writing proficiency as dependent variable, we found that only the CoV was a significant predictor, again being positive. The regression equation was significant (F(10, 11) = 6.05, p < 0.01), with an R^2 of 0.85. Details are shown in Table 5.

6. Discussion

The current study set out to investigate whether the finding by L&V that the degree of variability in L2 writing proficiency over time is better able to predict L2 writing proficiency gains than more traditional ID factors such as motivation and aptitude can be replicated in a different group of learners. However, it did so by improving the original study in several respects related both to the instruments measuring the ID factors and to the method of analysis. Also, because the L2 (English) learnt by the participants is far more distant to the native language of the learners in our sample (Chinese) than in the original study (Dutch), and there was only minimal extramural exposure to the L2 in our sample, we were able to control better for factors such as possible cross-linguistic transfer and the impact of L2 exposure.

Our results confirmed the original findings by showing that, first of all, none of the traditional ID factors—motivation, aptitude, or WM—was a significant predictor for final L2 writing proficiency, and secondly, that the degree of variability was significantly correlated with the L2 writing gains (RQ1). In addition, by administering additional regression analyses with both the ID factors and degree of variability as predictors, the current study found that variability was the best predictor both for final L2 writing proficiency (RQ2) and L2 writing proficiency gains (RQ3), irrespective of whether the participants were learning another language along with their L2 or not. A minor difference to the results of L & V is that the current study found starting L2 writing proficiency to be a significant predictor for final L2 writing proficiency.

Previous studies have shown that LA as measured by LLAMA had a meaningful relationship with L2 pronunciation, lexis and collocation knowledge among later learners (Granena & Long, 2012) and English proficiency scores operationalized as the official

Table 4
Multiple Linear Regression Analysis on the Final L2 Writing Proficiency.

Predictor variables	B	SE B	β
Significant predictors			
CoV	44.51	12.34	0.84
Starting L2 writing proficiency	0.73	0.19	0.90
Non-significant predictors			
Intrinsic motivation	−0.15	0.28	−0.13
Extrinsic motivation	−0.00	0.26	0.00
LLAMA-B	0.01	0.01	0.09
LLAMA-D	−0.00	0.02	−0.01
LLAMA-E	0.01	0.10	0.19
LLAMA-F	−0.01	0.10	−0.11
Operation span task	−0.03	0.02	−0.24
Language learning group	0.44	0.58	0.19
R^2	0.73		
F	2.92*		

Note. *, significant at 0.05 level.

Table 5
Multiple Linear Regression Analysis on the L2 Writing Proficiency Gains.

Predictor variables	B	SE B	β
Significant predictors			
CoV	44.51	12.34	0.63
Non-significant predictors			
Starting L2 writing proficiency	−0.27	0.19	−0.25
Intrinsic motivation	−0.15	0.28	−0.10
Extrinsic motivation	0.00	0.26	0.00
LLAMA-B	0.01	0.01	0.07
LLAMA-D	−0.00	0.02	−0.01
LLAMA-E	0.01	0.10	0.14
LLAMA-F	−0.01	0.01	−0.08
Operation span task	−0.03	0.02	−0.18
Language learning group	0.44	0.58	0.14
R^2	0.85		
F	6.05**		

Note. **, significant at 0.01 level.

school L2 proficiency tests (Artieda & Muñoz, 2016). However, the relationship between LA as measured by LLAMA with L2 writing gains has rarely been investigated. Being probably the first one to address this question, the current study found that vocabulary learning and sound-symbol corresponding ability had positive coefficients, while the other two (i.e. sound recognition and grammar inferencing ability) had negative coefficients in terms of the relationship with the final L2 writing proficiency as well as the gains. But none of these coefficients reached a significant level, which is comparable to the findings in L&V, in that general aptitude had a non-significant standard coefficient on predicting the final L2 writing proficiency.

WM as measured by the OSP task was previously found to play a role in different aspects of L2 processing. For example, WM was positively related to explicit knowledge of L2 grammar and vocabulary (Linck & Weiss, 2011, 2015). WM also interferes with the think-aloud protocol in learners' reading comprehension and written production, with those subjects with a higher WM capacity being bothered more with the think-aloud protocols (Goo, 2010). However, to our knowledge, the current study is the first attempt to examine the predictive power of WM as measured by OSP on L2 writing proficiency, and the finding did not provide any evidence for a direct significant impact from WM, similar to the findings regarding the LA.

Although it may be surprising to find that motivation did not play a role in predicting the final L2 writing proficiency in our study, we are not alone in this. For example, Lamb (2012) found a non-significant relationship between ideal L2 self and L2 proficiency among students from a provincial city and a rural area, compared to their peers residing in a metropolitan city. Similarly, Moskovsky et al. (2016) found that none of their four motivation factors (ideal L2 self, ought-to L2 self, language learning experience and intended effort to learning English) were significant, with all factors except ought-to L2 self showing a weak negative relation with L2 proficiency. They pointed out that, although motivation measures such as L2MSS are found to be good predictors of the intended effort to learn the language, motivation does not necessarily transfer to language learning behavior. Motivation may only have an impact on language learning when it actually translates to behavior, as shown in Kim and Kim (2011). This might also explain that, in our data, no aspect of motivation was a significant predictor for proficiency.

To conclude, the current study (as well as L&V) did not find that a high L2 motivation/high expectation of L2 self, high level of language aptitude, or high WM at the start of intensive language learning tends to lead to a higher level or higher gains of L2 writing proficiency later on. A fact that needs to be taken into consideration in this regard is the multi-faceted and dynamic nature of each ID factor (Dörnyei, 2010). Assuming a simple cause-effect relationship between the ID factors at one point and language learning success at a later point disregards the underlying dynamics, and a corresponding analysis is not really able to reveal the changes or the multi-level interactions within and between the ID factors themselves. Therefore, ID factors might be better accounted for in a complex dynamic perspective, under which the ID factors are understood as interconnected components of learner attributes which evolve over time, and interact with each other and the environment continuously (Dörnyei, 2010).

One way to incorporate the dynamic changes in development over time into the analysis is to include the variability in L2 writing proficiency scores over time as a factor in the analyses. In the present study, variability was found to correlate strongly and positively with the gains in L2 writing proficiency and to significantly predict both final L2 writing proficiency and gains in the regression analyses. These results are in line with the earlier findings in L&V, even though the current population was markedly different from theirs: while the participants in L&V were beginner learners of English with a starting age of 12–13 years and Dutch as their L1, the current participants were Chinese university students who had been learning English for several years already. This suggests that variability in L2 writing proficiency development might be a robust predictor for L2 writing proficiency (gains), not just in the populations studied so far.

To better illustrate how the degree of variability may predict the final writing proficiency and proficiency gains, we selected the learner with the highest degree of variability and the learner with the lowest degree of variability respectively, and visualized their development in L2 writing in a moving min-max graph (van Geert & van Dijk, 2002). The learner with the lowest degree of variability (Fig. 2) also had the lowest post score and a rather low gain score, while the learner with the highest variability (Fig. 3) had the highest post score, as well as the highest gain score.

The moving min-max adopts a moving window of three data points, and the bandwidth between the minimum and maximum line

shows the degree of variability. The larger the space, the larger the variability. Periods of high variability indicate that the learner is varying his or her behavior a lot in a short period of time and may point to the learner trying out new things or maybe overusing certain aspects of language knowledge or skills. During such a phase of high variability, the system is often unstable and is possibly moving from one phase to another (de Bot et al., 2007; Spoelman & Verspoor, 2010). As Fig. 2 reveals, student A went through one such period, from the 5th to the 9th text. After the 9th text, s/he entered into a more stable phase, and displayed a slightly higher writing proficiency compared to the starting period. Fig. 3 demonstrates the developmental trajectory for student B, who had the highest degree of variability during the academic year. The graph reveals that student B showed a larger degree of variability than student A during the whole academic year, veering between higher and lower proficiency scores, but still exhibiting a clear upward trend.

In order to evaluate this finding, the role of variability needs to be understood more thoroughly. In our analyses, variability was used in regression analyses as a predictor for the final L2 writing proficiency and L2 writing proficiency gains, in the same way as traditional ID factors. However, technically, variability is not an ID factor and is distinct from the traditional ID factors, e.g., motivation, LA, WM, learning style, or personality. Instead, we would argue that variability can better be understood as a reflection of the changes and the development occurring within the system (the learner), and these changes in turn are caused by the dynamics and the interconnectedness of the ID factors and other factors involved in language learning. In other words, a higher variability in language proficiency over time does not *cause* higher language proficiency or higher gains in itself, but is instead the *surface reflection* of the dynamic system of a language learner, who is more likely to gain more in proficiency and end up with a higher proficiency – and that is why a higher variability may predict these changes, too.

To view variability in such a way is in line with a CDST perspective on language development: the learner and the language learned are regarded as complex, dynamic systems which in turn consist of dynamic and interrelated subsystems (Verspoor et al., 2008). Examples of these subsystems may include the learner's personality, their desire to learn the language (e.g. motivation and attitude), their cognitive capacities (e.g. LA and WM), the complexity of the linguistic aspect currently learnt, or the input and instruction the learner receives. As Dörnyei (2010) suggested, these aspects all change over time and are interconnected with each other:

> They (individual differences) are not at all stable but show salient temporal and situational variation, and they are not monolithic either but constitute complex constellations that are made up of different parts that interact with each other and the environment synchronically and diachronically. (pp. 440)

One consequence of this interrelatedness of the multi-faceted, changing factors is that the learner will show variability in their learning process in ways that cannot be fully predicted. Such variability is typical of complex dynamic systems and their self-organization. To illustrate this with a hypothetical example, a learner's higher WM or better language aptitude may make their language learning experience less demanding and thus more enjoyable and rewarding, and this in turn may motivate the learner - and the highly motivated intensive language learning in return may consolidate the cognitive abilities. The fact that IDs change over time was found in a related study with the same population as in the current study. The learners improved both in language aptitude (sound recognition ability and sound-symbol corresponding ability) and WM during the same period of observation. This change was related to the intensity of language learning (with the E/R learners gaining more), and corroborates an understanding of ID factors in language learning as changing and interrelated. In this sense, it seems less surprising that the levels of motivation, language aptitude, and WM measured once at the start of the academic year were not able to reliably predict the outcome of the learning trajectory at the end of the year in the current study.

If variability in proficiency scores is understood as a product of the dynamics of the language learning process, what might variability point to in the learner and learner-based factors? With this question in mind, we carried out a correlation analysis trying to identify a possible relationship between CoV and the motivational factors. The results showed that two factors, i.e. family influence and ought-to L2 self, were significantly correlated with the CoV. See Table 6 for details.

As pointed out by Magid (2009), family influence plays an important role on the Ought-to L2 self among Chinese learners, and has a powerful impact on Chinese learners' motivation to learn English. Ought-to L2 self refers to the attributes that one supposes one ought to have (Dörnyei, 2010), and it is often a reflection of the expectations from the outside world for the learner. As proposed by Chen, Warden, and Chang (2005), for Chinese foreign language learners, there might be a culturally specific motivator called the "Chinese Imperative" (p.623), which is a reflection of the emphasis on "requirements" from the family and society, and the emphasis on exam
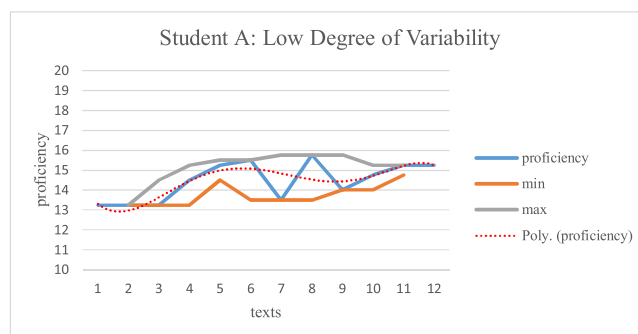

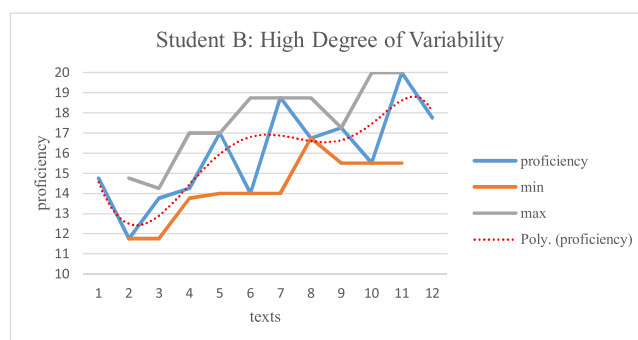**Fig. 2.** The Moving Min-max Graph of the Learner with Lowest Degree of Variability.

Fig. 3. The Moving Min-max Graph of the Learner with Highest Degree of Variability.

Table 6
Significant Correlations between CoV and Two Motivational Factors.

Pearson	r	p (two-tailed)
Family influence & CoV	0.454*	0.034
Ought-to L2 self & CoV	0.478*	0.024

Note. *, correlation is significant at 0.05 level.

results. Perhaps such learners are disappointed with themselves when they score relatively low at one time and work really hard to do better the next time, leading to more variability. All in all, the fact that the CoV was significantly and positively related to these two factors, may indicate that learners under more pressure from the outside world are more willing to try and reach high, and tend to show more variability in L2 writing. Therefore, apart from purely being a result of the system's self-organization, we would like to suggest tentatively that variability in language learning might also in part be seen as a symptom of a learner's higher willingness to seek their limits in language learning, a propensity to be more daring and adventurous, to invest more in the learning process and therefore both reach higher peaks at some point, but also fall back to a lower level on another occasion. As Lowie and Verspoor (2019) phrased it, "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).

To conclude, instead of being a direct reason or cause of final L2 writing proficiency or gains, variability is a symptom of the dynamic changes and the interconnectedness of the factors in language learning. But more variability is associated with systems that tend to develop to a higher level, and thus with learners who tend to end up with a higher (gain in) (writing) proficiency and are able to use the factors involved in language learning to their advantage. This might be the reason that variability was found to be the best (and only) predictor for our learners' L2 attainment (as reflected in their writings).

The fact that, variability proved to be such a powerful predictor for proficiency development both in the current study and in L&V and in such different settings, confirms the argument of pioneering CDST studies (e.g. de Bot et al., 2007; Lowie & Verspoor, 2015; Spoelman & Verspoor, 2010) to regard variability as an important source to investigate language development.

This finding suggests that teachers should realize the nonlinearity of language learning. While learning a new language, learners may reach a new maximum at one time and then regress the next time, which, however, does not necessarily point to a negative destabilization. As Lowie (2013) put it, "where there is variability, there is development" (p. 21). Language teachers should be aware of the value of such variability. But even more importantly, these findings suggest that important grades (e.g. the final grade for a class) should not be based on a single performance as especially learners who are in a phase in which they are developing fast in a (sub) skill may vary quite a bit from one performance to the next.

7. Conclusion and limitations

The current study was a replication of L&V and tried to explore the best predictors for final L2 writing proficiency and L2 writing proficiency gains. In line with the findings in L&V, no traditional ID factor was found to be a significant predictor for final L2 writing proficiency, while variability in L2 writing proficiency development was significantly correlated with higher L2 writing proficiency gains. In addition, two multiple linear regression analyses showed that variability can predict final L2 writing proficiency and gains when the starting proficiency was controlled, while traditional ID factors fail to do so. This led to the conclusion that, firstly, ID factors when measured at one point of time may not always be able to reliably predict L2 attainment or gains at a later point in time, as the ID factors are multi-faceted and dynamic in nature; secondly, that variability is a result of the dynamic changes and interconnectedness of the internal and external factors in language learning and more variability seems to be associated with more successful learners.

However, while this association has now been detected in two rather distinct populations of learners, further research is needed to corroborate this finding and explore it further. Given this dynamic nature of language development, a drawback of this study is that the ID factors were not measured longitudinally in the way L2 writing proficiency was. To fully understand the development and the

factors impacting the development at the individual level, not only the linguistic aspects should be measured in a longitudinal dimension, but we suggest that future studies should also trace the dynamics and interconnectedness of the ID factors over time. In addition, the motivation questionnaire used in the current study is a general English learning motivation questionnaire, which is comprehensive but may still be too general and unable to capture task-specific aspects such as the desire to improve writing skills, or the attitude towards each writing topic. Future studies are encouraged to use more L2 writing-specific motivation questionnaires. Another limitation of the current study concerns the relatively small sample size, and this may warrant caution in interpretation of the results of the factor analysis and linear regression. Future studies corroborating the findings in the current study are suggested to study a larger population.

Funding

Declaration of Competing Interest

The authors report no declarations of interest.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition, 30*(4), 481–509. https://doi.org/10.1017/S027226310808073X.

Artieda, G., & Muñoz, C. (2016). The LLAMA tests and the underlying structure of language aptitude at two levels of foreign language proficiency. *Learning and Individual Differences, 50*, 42–48. https://doi.org/10.1016/j.lindif.2016.06.023.

Atkins, P. W. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics, 19*(4), 537–552. https://doi.org/10.1017/s0142716400010353.

Baddeley, A. D. (1992). Working memory. *Science, 255*(5044), 556–559. https://doi.org/10.4324/9781912282418.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2.

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (pp. 47–90). New York: Academic Press.

Bergsleithner, J. M. (2010). Working memory capacity and L2 writing performance. *Ciências & Cognição, 15*, 2–20.

Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. *Critical Care, 7*(6), 451–459. https://doi.org/10.1186/cc2401.

Cancino, H., Rosansky, E., & Schumann, J. H. (1978). The acquisition of English negatives and interrogatives by native Spanish speakers. In E. M. Hatch (Ed.), *Second language acquisition: A book of readings*. Rowley, Mass: Newbury House Publishers.

Carroll, J. B. (1965). The prediction of success in foreign language training. In R. Glaser (Ed.), *Training, research, and education* (pp. 87–136). New York: Wiley.

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118).

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test*. San Antonio, TX, US: Psychological Corporation.

Chen, J. F., Warden, C. A., & Chang, H. T. (2005). Motivators that do not motivate: The case of Chinese EFL learners and the influence of culture on motivation. *TESOL Quarterly, 39*(4), 609. https://doi.org/10.2307/3588524.

de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic view as a complementary perspective. *Bilingualism: Language and Cognition, 10*(01), 51–55. https://doi.org/10.1017/S1366728906002811.

de Weerth, C., van Geert, P., & Hoijtink, H. (1999). Intraindividual variability in infant behavior. *Developmental Psychology, 35*(4), 1102–1112. https://doi.org/10.1037/0012-1649.35.4.1102.

Denhovska, N., Serratrice, L., & Payne, J. (2016). Acquisition of second language grammar under incidental learning conditions: The role of frequency and working memory. *Language Learning, 66*(1), 159–190. https://doi.org/10.1111/lang.12142.

Dörnyei, Z. (2000). Motivation in action: Towards a process-oriented conceptualisation of student motivation. *British Journal of Educational Psychology, 70*, 519–538. https://doi.org/10.1348/000709900158281.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review, 19*, 42–68. https://doi.org/10.1075/aila.19.05dor.

Dörnyei, Z. (2010). The relationship between language aptitude and language learning motivation: Individual differences from a dynamic system perspective. In E. Macaro (Ed.), *The bloomsbury companion to second language acquisition* (pp. 247–267). London: Bloomsbury Publishing Plc.

Dörnyei, Z. (2009a). Individual differences: Interplay of learner characteristics and learning environment. *Language Learning, 59*(SUPPL. 1), 230–248. https://doi.org/10.1111/j.1467-9922.2009.00542.x.

Dörnyei, Z. (2009b). The L2 motivational self system. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). https://doi.org/10.21832/9781847691293-003.

Dörnyei, Z., & Csizér, K. (2012). How to design and analyze surveys in second language acquisition research. In A. Mackey, & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 74–94). https://doi.org/10.1002/9781444347340.ch5.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Malden, MA: Blackwell Publishing Ltd.

Dörnyei, Z., & Ushioda, E. (Eds.). (2009). *Motivation, language identity and the L2 self*. Multilingual Matters.

Goo, J. (2010). Working memory and reactivity. *Language Learning, 60*(4), 712–752. https://doi.org/10.1111/j.1467-9922.2010.00573.x.

Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA language aptitude test. In G. Granena, & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105–129). Amsterdam: John Benjamins Publishing Co.

Granena, G., & Long, M. H. (2012). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research, 29*(3), 311–343. https://doi.org/10.1177/0267658312461497.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition, 19*(3), 379–400.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*(01), 25. https://doi.org/10.1017/S0272263100010457.

Hernández, T. A. (2010). The relationship among motivation, interaction, and the development of second language oral proficiency in a study-abroad context. *Modern Language Journal, 94*(4), 600–617. https://doi.org/10.1111/j.1540-4781.2010.01053.x.

Hou, J., Verspoor, M., & Loerts, H. (2016). An exploratory study into the dynamics of Chinese L2 writing development. *Dutch Journal of Applied Linguistics, 5*(1), 65–96. https://doi.org/10.1075/dujal.5.1.04loe.

Huang, T., Loerts, H., & Steinkrauss, R. (2020). The impact of second- and third-language learning on language aptitude and working memory. *International Journal of Bilingual Education and Bilingualism, 0*(0), 1–17. https://doi.org/10.1080/13670050.2019.1703894.

Huang, T., Steinkrauss, R., & Verspoor, M. (2020). Learning an L2 and L3 at the same time : Help or hinder? *International Journal of Multilingualism, 0*(0), 1–17. https://doi.org/10.1080/14790718.2020.1779726.

Kim, Y.-K., & Kim, T.-Y. (2011). The effect of Korean secondary school students' perceptual learning styles and ideal L2 self on motivated L2 behavior and English proficiency. *Korean Journal of English Language and Linguistics, 11*(1), 21–42. https://doi.org/10.15738/kjell.11.1.201103.21.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11*(02), 261–271. https://doi.org/10.1017/S1366728908003416.

Lalonde, R. N., & Gardner, R. C. (1985). On the predictive validity of the attitude/motivation test battery. *Journal of Multilingual and Multicultural Development, 6*(5), 403–412. https://doi.org/10.1080/01434632.1985.9994214.

Lamb, M. (2012). A self system perspective on young adolescents' motivation to learn English in urban and rural settings. *Language Learning, 62*(4), 997–1023. https://doi.org/10.1111/j.1467-9922.2012.00719.x.

Li, S. (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics, 36*(3), 385–408. https://doi.org/10.1093/applin/amu054.

Li, S. (2016). The construct validity of language aptitude. *Studies in Second Language Acquisition, 38*(4), 801–842. https://doi.org/10.1017/S027226311500042X.

Linck, J. A., & Weiss, D. J. (2011). Working memory predicts the acquisition of explicit L2 knowledge. In C. Sanz, & R. P. Leow (Eds.), *Implicit and explicit language learning: Cognition, processes, and knowledge in SLA and bilingualism* (pp. 101–114). https://doi.org/10.1017/CBO9781107415324.004.

Linck, J. A., & Weiss, D. J. (2015). Can working memory and inhibitory control predict second language learning in the classroom? *Sage Open, 5*(4), 1–11. https://doi.org/10.1177/2158244015607352.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21*, 861–883. https://doi.org/10.3758/s13423-013-0565-2 (1531-5320 (Electronic)).

Lowie, W. (2013). The CEFR and the dynamics of second language learning: Trends and challenges. *Language Learning in Higher Education, 2*(1), 17–34.

Lowie, W., & Verspoor, M. (2015). Variability and variation in second language acquisition orders: A dynamic reevaluation. *Language Learning, 65*(1), 63–88. https://doi.org/10.1111/lang.12093.

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

Mackey, A., Adams, R., & Stafford, C. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning, 60*(3), 501–533. https://doi.org/10.1111/j.1467-9922.2010.00565.x.

Magid, M. (2009). The L2 motivational self system from a Chinese perspective: A mixed methods study. *Journal of Applied Linguistics, 6*(1), 69–90. https://doi.org/10.1558/japl.v6i1.69.

Meara, P. (2005). *LLAMA language aptitude tests: The manual*. Swansea: Lognostics.

Mitchell, A. E., Jarvis, S., O'Malley, M., & Konstantinova, I. (2015). Working memory measures and L2 proficiency. In Z. Wen, M. Borges, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 270–284). Bristol, UK: Multilingual Matters.

Miyake, A., & Friedman, N. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy, & L. E. J. Bourne (Eds.), *Foreign language learning: Psyhcolinguistic studies on training and retention* (pp. 339–364).

Moskovsky, C., Assulaimani, T., Racheva, S., & Harkins, J. (2016). The L2 motivational self system and L2 achievement: A study of Saudi EFL learners. *Modern Language Journal, 100*(3), 641–654. https://doi.org/10.1111/modl.12340.

Payne, J. S., & Ross, B. M. (2005). Synchronous CMC, working memory, and L2 oral proficiency development. *Language Learning & Technology, 19*(3), 35–54.

Rai, M. K., Loschky, L. C., Harris, R. J., Peck, N. R., & Cook, L. G. (2011). Effects of stress and working memory capacity on foreign language readers' inferential processing during comprehension. *Language Learning, 61*(1), 187–218. https://doi.org/10.1111/j.1467-9922.2010.00592.x.

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research, 17*(4), 368–392. https://doi.org/10.1177/026765830101700405.

Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA: A replication and extension of Reber, Walkenfield and Hernstadt. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211–266). Amsterdam: John Benjamins Publishing Co.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics, 25*(2005), 46–73. https://doi.org/10.1017/S0267190505000036.

Rogers, V., Meara, P., Barnett-legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *JESLA, 1*, 49–60.

Saito, K., Dewaele, J.-M., & Hanzawa, K. (2017). A longitudinal investigation of the relationship between motivation and late second language speech learning in classroom settings. *Language and Speech, 60*(4), 614–632. https://doi.org/10.1177/0023830916687793.

Schepens, J. J., van der Slik, F., & van Hout, R. (2016). L1 and L2 distance effects in learning L3 Dutch. *Language Learning, 66*(1), 224–256. https://doi.org/10.1111/lang.12150.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide*. Pittsburgh: Psychology Software Tools Inc.

Serafini, E. J., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition, 38*, 607–646. https://doi.org/10.1017/S0272263115000327.

Skehan, P. (1991). Individual differences in second language learning. *Studies in Second Language Acquisition, 13*(2), 275–298. https://doi.org/10.1017/S0272263100009979.

Skehan, P. (2012). Language aptitude. In S. M. Gass, & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 381–395). Oxon: Routledge.

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2009). Long-term relationships among early first language skills, second language aptitude, second language affect, and later second language proficiency. *Applied Psycholinguistics, 30*(4), 725–755. https://doi.org/10.1017/S0142716409990099.

Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of finish. *Applied Linguistics, 31*(4), 532–553. https://doi.org/10.1093/applin/amq001.

Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese and Iranian learners of English: A comparative study. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 66–97). https://doi.org/10.21832/9781847691293-005.

Thelen, E., & Smith, L. B. (1994). In E. Thelen, & L. B. Smith (Eds.), *A dynamic systems approach to the development of cognition and action*. Cambridge, Mass: MIT Press.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505. https://doi.org/10.3758/BF03192720.

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

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. https://doi.org/10.1111/j.1540-4781.2008.00715.x.

Waller, L., & Papi, M. (2017). Motivation and feedback: How implicit theories of intelligence predict L2 writers' motivation and feedback orientation. *Journal of Second Language Writing, 35*, 54–65. https://doi.org/10.1016/j.jslw.2017.01.004.

(Edward) Wen, Z. (2015). Working memory in second language acquisition and processing: The phonological/executive model. In Z. Wen, B. M. Mailce, & M. Arthur (Eds.), *Working memory in second language acquisition and processing* (pp. 41–62). https://doi.org/10.21832/9781783093595.

(Edward) Wen, Z. (2016). *Working memory and second language learning: Towards an integrated approach*. Bristol: Multilingual Matters.

(Edward) Wen, Z. (2019). Working memory as language aptitude: The phonological/executive model. In Z. Wen, P. Skehan, A. Biedroń, S. Li, & R. L. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 187–214). New York: Routledge.

Williams, J. N. (2012). Working memory and SLA. In S. M. Gass, & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 427–441). Oxon: Routledge.

Yang, L., Xu, K., & Lu, P. (Eds.). (2010). *Contemporary college English* (second edition). Beijing: Foreign Language Teaching and Research Press.

Ting Huang recently gained her PhD degree at University of Groningen. She studies second language development.

Rasmus Steinkrauss is an assistant professor at University of Groningen. His expertise is in L1 and L2 development, and Usage-based/Cognitive Linguistics.

Marjolijn Verspoor is an interim professor of language learning at University of Groningen and a professor in Applied Linguistics at University of Pannonia. Her expertise is in English language and second language acquisition