Individualization of practice distribution in second language grammar learning

A role of metalinguistic rule rehearsal ability and working memory capacity

Yuichi Suzuki
Kanagawa University

The aim of the present study was establishing to what extent individual differences in cognitive aptitudes were associated with second language (L2) morphological acquisition under different practice distribution. Sixty participants studied morphological rules of a novel miniature-language system in order to use them for oral production. They engaged in four training sessions in either shorter-spaced learning (twice a week) or longer-spaced learning conditions (once a week). Their oral production performance both during and after the training was related to their metalinguistic rule rehearsal ability (MRRA) and working memory capacity (WMC). Multiple regression analyses revealed that MRRA predicted learners' training and posttest performance mainly under the longer-spaced condition, while WMC played a limited role at best under both learning conditions. These results suggest that practice distribution may be individualized based on learners' aptitude strengths to optimize L2 morphological learning.

Keywords: aptitude-treatment interaction, aptitude complex, metalinguistic rule rehearsal ability, working memory, practice distribution, grammar learning, oral production

When teaching second language (L2), it is essential to determine the most optimal schedule for multiple learning sessions aimed at facilitating learning and retention of L2 skills. This issue has been a subject of growing body of research aiming to ascertain whether changing the timing of study sessions, while keeping the total amount of learning time constant, influenced the retention of L2 knowledge and skills (Bird, 2010; Bloom & Shuell, 1981; Miles, 2014; Nakata, 2015; Rogers, 2015; Serrano & Huang, 2018; Suzuki, 2017; Suzuki & DeKeyser, 2017a). These studies demonstrate that timing does influence L2 acquisition, corroborating the

Appendixes available from https://doi.org/10.1075/jsls.18023.suz.additional
https://doi.org/10.1075/jsls.18023.suz
Journal of Second Language Studies 2:2 (2019), pp. 170–196. issn 2542-3835 | e-issn 2542-3843

findings yielded by a large body of cognitive psychology research (see Carpenter et al., 2012; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; and Toppino & Gerbier, 2014 for review). Yet, to date, whether optimal levels of learning distribution vary depending on individual differences in learners' characteristics such as cognitive aptitudes has not been examined. The only exception is an exploratory study conducted by Suzuki and DeKeyser (2017b), which is extended in the current work to examine the role of cognitive aptitudes on the effectiveness of shorter-spaced and longer-spaced learning conditions for L2 morphological structures.

Research on aptitude for L2 learning has been conducted for several decades (see a historical overview provided by Carroll, 1981), and recent progress in the field involves conceptualizing and developing new measures for cognitive aptitude components (Doughty, in press; Granena, 2013, 2016; Linck et al., 2013; Robinson, 2007; Skehan, 2002). In extant research, individual differences in cognitive aptitude were found to moderate the effectiveness of different types of L2 instruction (see Li, 2015; and Vatz, Tare, Jackson, & Doughty, 2013 for review). In other words, one type of instruction is more or less effective than another depending on the individual's aptitude strengths and weaknesses; this is known as aptitude-treatment interaction (ATI) (Cronbach & Snow, 1977). Examining ATI is important for both theoretical and practical reasons. From a practical point of view, ATI patterns may potentially reveal the optimal matching of instruction types to the learners' aptitude profiles (Wesche, 1981). Seen from a more theoretical angle, they also make it possible to infer the effect of different treatments on employment of different learning processes. For instance, if an aptitude for linguistic analysis (e.g., grammatical sensitivity) was related to Treatment A but no to Treatment B, we can infer that Treatment A induced learning processes that necessitated analysis of linguistic structure more strongly (DeKeyser, 2012).

The current study aims to reveal L2 morphological learning processes, in the context of explicit grammar instruction, under shorter-spaced and longer-spaced learning conditions in the ATI framework (Cronbach & Snow, 1977; Vatz et al., 2013). The aim is to advance the current understanding of ATI for L2 learning in two significant aspects. First, instead of using an isolated aptitude component, the author investigated the extent to which a meaningful set of aptitudes or *aptitude complexes* (Robinson, 2007; Snow, 1987) can predict L2 acquisition. Despite the evident potential of assembling multiple aptitude components, there is paucity of research in which the usefulness of different sets of aptitude complexes is compared (Dörnyei & Ryan, 2015). Secondly, the focus of the current study is on the role of aptitude in different stages of L2 learning (Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong, 2014; Skehan, 2002, 2016; Ullman, 2015). As L2 acquisition is contingent on multiple and complex processes, tracking the learning processes during the training sessions can allow for a more fine-grained analysis of ATI patterns. As a part of this investigation, the effects of aptitudes were examined over four one-hour training sessions as well as via two post-test performance scores in order to track the changes of aptitude at different time points.

The effects of shorter-spaced and longer-spaced learning

Findings from a large body of research in cognitive psychology indicate that different distribution influences learning and retention of knowledge and skills (see Carpenter et al., 2012; Cepeda et al., 2006; Toppino & Gerbier, 2014 for review). Research compares different intersession intervals (ISIs) in combination with different retention intervals (RIs). For instance, shorter-spaced learning (e.g., 3-day ISI) and longer-spaced learning (e.g., 14-day ISI) conditions on retention of memory with different delays (e.g., 7-day and 35-day RIs). A large-scale study conducted by Cepeda, Vul, Rohrer, Wixted, and Pashler (2008) demonstrated that the optimal ISI is 8–43% of RI for recall memory and 7–24% of RI for recognition memory, suggesting that the ratio of ISI-RI informs how ISI can be chosen optimally. For best recall performance, for example, 3-day ISI was optimal for the 7-day RI (ISI-RI ratio of 43%), while for the 35-day RI, the 8-day ISI was the best (ISI-RI ratio of 23%).

Following upon the Cepeda et al.'s (2008) framework, an emerging line of research has examined the effects of ISI on L2 grammar learning with different RIs (Bird, 2010; Rogers, 2015; Suzuki, 2017; Suzuki & DeKeyser, 2017a). The earlier studies (Bird, 2010; Rogers, 2015) found that the optimal ISI-RI ratio for L2 grammar learning is similar to the range of ISI-RI ratios reported by Cepeda et al. (2008) based on their experiment, suggesting the potential for generalizing the cognitive psychology findings to L2 grammar learning. More recent studies (Suzuki, 2017; Suzuki & DeKeyser, 2017a), however, cast doubt on the earlier results, as they demonstrate that shorter-spaced learning may be more beneficial than longer-spaced learning for proceduralization of L2 grammatical knowledge (e.g., the acquisition of oral production skills). Although consensus on this issue is presently lacking (see Suzuki, 2017 for a detailed discussion), available evidence clearly indicates that different levels of learning distribution impact L2 grammar acquisition. This line of investigation is refined for the current study; the primary focus of the current study is to examine the extent to which internal factors (i.e., cognitive aptitudes) moderate the effectiveness of different practice distribution.

Aptitude complexes and L2 grammar learning

Many researchers have investigated to what extent one aptitude component moderates the effectiveness of different L2 learning conditions (e.g., Goo, 2012; Sanz, Lin, Lado, Stafford, & Bowden, 2014; Sheen, 2007). More recently, however, authors of L2 studies have started to investigate how different sets of aptitudes (aptitude complexes) interact with learning conditions (Erlam, 2005; Li, 2013; Morgan-Short et al., 2014; Yilmaz, 2013). Examining the role of multiple aptitude components, rather than an isolated aptitude component, is informative as it can potentially reveal specific L2 learning processes. L2 acquisition involves complex learning processes, which are facilitated or inhibited by different aptitude types (DeKeyser & Koeth, 2010; Linck et al., 2013).

The objective of the current study is to examine how different aptitude complexes interact with different levels of L2 learning distribution. To the best of the author's knowledge, Suzuki and DeKeyser (2017b) were the only researchers to examine the role of aptitude complexes for longer-spaced/shorter-spaced L2 learning. They examined the role of two cognitive aptitudes in L2 morphological learning under two different learning intervals. The first aptitude component was language analytic ability (LAA), defined as the ability "to infer rules of language and make linguistic generalizations or extrapolations" (Skehan, 1998, p. 204), and the second aptitude component was working memory capacity (WMC), operationalized as a control mechanism that stores and manipulates incoming information until it is integrated into the cognitive process (Linck, Osthus, Koeth, & Bunting, 2014). LAA was assessed by the grammatical-inferencing sub-test of the LLAMA aptitude test (Meara, 2005) in which the participants studied the grammar of an unfamiliar language for five minutes before testing their learning rates. WMC was measured by the operation-span (Ospan) task in which the participants performed the dual task of remembering a list of alphabet letters while solving math problems. These aptitude scores were compared with the participants' performance on the outcome tests (immediate and two delayed posttests).

The results yielded by these analyses showed that LAA was related to the learning outcomes of the longer-spaced learning group only ($.66 < r < .72$), while WMC was related to learning outcomes among the shorter-spaced learning group only ($.41 < r < .56$). Based on the ATI pattern found, Suzuki and DeKeyser (2017b) posited that similar morphological rules caused strong interference in their memorization, particularly under shorter-spaced learning condition (with 1-day interval between sessions), and higher WMC facilitated learners' discrimination of the similar morphological rules. For the association of LAA with longer-spaced learning, higher LAA might have facilitated a deeper understanding of language structure, allowing learners to retain their knowledge over a longer (7-day) interval. In sum, a combination of LAA and WMC *differentially* predicted learning outcomes under different levels of learning distribution.

Although Suzuki and DeKeyser (2017b) acknowledged the exploratory nature of their study (given that the number of participants was small, $n = 40$), their findings offer a promising direction for further research aimed at gaining a better understanding of the role of aptitude complexes in longer-spaced/shorter-spaced learning. The goal of the present study is to extend Suzuki and DeKeyser's work by examining another aptitude component (rote memory ability, RMA) in addition to LAA and WMC. While Suzuki and DeKeyser (2017b) chose a combination of LAA and WMC as an aptitude complex, adding another aptitude complex may be even more useful for differentially predicting shorter-spaced and longer-spaced learning outcomes. Robinson (2007) presented an aptitude complex for explicit rule learning, which is relevant for the current study. It consists of metalinguistic rule rehearsal ability (MRRA) and memory for contingent text (MCT). MRRA is segregated into LAA and RMA, while MCT refers to WMC.[1] According to Robinson (2007), WMC is relevant for L2 explicit learning because, when learners apply grammatical rules, they need to maintain and rehearse the rules in their WM for subsequent L2 comprehension and production. In addition, MRRA was a strong predictor of the learning outcomes in the rule-instruction learning condition (Robinson, 1997), supporting the conceptualization of the aptitude complex for explicit rule learning. Although Suzuki and DeKeyser (2017b) utilized LAA only, integrating RMA with LAA (i.e., MRRA) can serve as a more powerful predictor of learning rates ("ability factor" in Robinson, 2007), because L2 morphological learning relies on vocabulary or paired-associate learning (Ellis & Schmidt, 1997; Suzuki & DeKeyser, 2017a). In sum, in the current study, MRRA and WMC are examined as predictors of acquisition of L2 morphological features under shorter-spaced and longer-spaced learning conditions.

Aptitude complex for different L2 learning stages

Aptitude may play selective roles in different stages of L2 learning (Morgan-Short et al., 2014; Serafini & Sanz, 2016; Skehan, 2002). Skehan (2002, 2016) posited that a different set of aptitudes may relate to specific L2 developmental stages. He argues that phonemic coding ability and WMC are particularly important in the earlier stage (noticing form in input). In the later stage, LAA is needed

1. Robinson stipulated two separate components of WM: WM for text and speed of WM for text. Since no measure for speed of WM was provided in his model, it was simplified to include a single component for WM.

more strongly for identifying and extracting linguistic patterns from the noticed input, while various facets of memory may be implicated in faster, more efficient retrieval of learned rules (proceduralization). Despite its evident potential, a majority of studies focuses on the relationship between aptitude and learning outcomes (after the treatment). Among the limited number of extant empirical studies, the work of Morgan-Short et al. (2014), which draws on Ullman's declarative-procedural model (Ullman, 2015), is particularly relevant for the present investigation. These authors showed that declarative memory ability, typically operationalized as RMA, predicted the earlier stage of artificial grammar learning (after completing two training sessions), whereas procedural memory ability, commonly operationalized as sequence learning ability, predicted the later stage of acquisition (after completing four sessions). Examining L2 acquisition during the training phase may thus offer more fine-grained analysis of L2 learning processes.

In the current study, the training duration was extended to four one-hour sessions delivered across weeks, and the two delayed post-tests were administered one and four weeks after completing the last training session. This arrangement allowed us to examine how the contribution of aptitudes varies at different time points. Because of their particular relevance for shorter-spaced and longer-spaced learning, two fundamental learning stages – learning versus forgetting/retention (Anderson, 2015) – are distinguished in the current study. In this context, the learning phase concerns learning gains during training sessions, whereas the knowledge retention and forgetting phase pertains to the period during which no re-learning opportunities are provided. In order to examine these two systems separately, as a part of the current study, nine assessment tests were administered during and after the training sessions. These tests were taken by the study participants at the end of Sessions 1–4 (learning phase) and at the beginning of Sessions 2–4 (forgetting/retention phase), along with two delayed post-tests administered during Sessions 5 and 6 (see Research Design and Procedure). This research design thus allows for delineation of the (a) learning and (b) retention/forgetting phases of L2 grammar learning, in addition to making the distinction between the earlier and the later stages of learning.

The present study

The present study was an extension of the ATI study conducted by Suzuki and DeKeyser (2017b), as the aim was to examine the role of different cognitive aptitudes in the acquisition of L2 morphology under the shorter-spaced and longer-spaced learning conditions. The data collected for a larger project of distributed grammar practice (see Suzuki, 2017 for comprehensive results of the relative effects of shorter- and longer-spaced learning conditions) were re-analyzed for probing the roles of cognitive aptitudes. In Suzuki's study (2017), sixty participants were trained on an element of the miniature language system loosely based on Spanish (i.e., present-progressive morphological markers). They all completed four one-hour training sessions during which they learned six morphological rules accompanied by new vocabulary and grammatical explanation. However, half of the study sample took part in training sessions once a week (7-day intervals, longer-spaced learning group, $n = 30$), whereas the other half received the same training sessions twice a week (3.3-day interval, shorter-spaced learning group, $n = 30$). The shorter-spaced learning interval thus differed from the one-day interval employed by Suzuki and DeKeyser (2017b) because it was not feasible to recruit participants with that schedule. This change provided a stricter test case for the generalizability of Suzuki and DeKeyser's (2017b) ATI findings. Four aptitude tests were administered to assess LAA, WMC, and RMA, which were assembled into MRRA and WMC. The present study was guided by the following research questions (RQs):

1. To what extent do MRRA and WMC predict L2 acquisition of morphological rules under the shorter-spaced and longer-spaced learning conditions?
2. Does the contribution of the aptitudes vary over time?
   a. Does the contribution of the aptitudes differ between the earlier and the later learning phases?
   b. Does the contribution of the aptitudes differ between the learning and the retention phases?

For RQ1, it was predicted that MRRA would be exclusively related to the learning outcomes of the longer-spaced learning group, whereas WMC would be exclusively related to those pertaining to the shorter-spaced learning group (Suzuki & DeKeyser, 2017b). It is important to note that the relationship with the aptitudes might differ from Suzuki and DeKeyser's (2017b) findings, in the shorter-spaced learning group in particular, as the learning interval was operationalized differently in the study (3.3-day interval as opposed to 1-day interval).

The second ~~research question~~ was exploratory and was addressed to examine whether the role of aptitudes would change at different time points (Skehan, 2002). As delineated above, learning and forgetting/retention phases were distinguished and different predictions were given for each phase. For learning gains during training sessions, it was postulated that the contribution of aptitudes would follow the expected pattern (e.g., longer-spaced learning -> MRRA, shorter-spaced learning -> WMC). For the forgetting/retention phase, however, the contribution of both aptitudes was expected to be greater in the longer-spaced learning group than in the shorter-spaced learning group. This difference was anticipated because longer intervals should lead to more extensive forgetting, while also revealing individual differences in aptitudes.

Methods

Participants

Sixty students at a private Japanese university (of whom 20 were males and 40 females) participated in the study. Their mean age was 19.63 ($SD = .96$). None of the participants had learned Spanish before. Twenty-eight participants studied a foreign language other than English at the university (15 and 13 participants in the shorter-spaced and longer-spaced groups, respectively). No significant differences were found in the mean length of study of these foreign languages (in months) between the shorter-spaced and the longer-spaced learning groups ($M = 23.00$, 21.04; $SD = 20.7$, 13.15), $t\ (26) = 0.29$, $p = .77$.

Instruments

Target structures

A miniature language called "Supurango" loosely based on Spanish, was created for this study. Spanish was chosen because the phonology is easier to learn for Japanese speakers, allowing for investigating the acquisition of grammar. The participants in Suzuki and DeKeyser's (2017b) study had some prior knowledge of the target structure (*-te* form) before the experiment, while the students that took part in the present study were trained on morphological structures in a novel miniature language, controlling for the extraneous factors (i.e., no prior knowledge or no exposure beyond the laboratory setting).

The target grammatical structure was present progressive (PP), which was expressed by a morphological marking on a verb. As shown in Table 1, the language had six morphological rules depending on the verb ending. While three simple verb types (AR, ER, IR) required a change in the verb ending only, the other three complex verb types (AS, ES, IS) involved two transformations in the first vowel and verb ending.[2]

Four action verbs were chosen for each verb type, resulting in a set of 24 verbs used in training (see Appendix S1 in the online supporting information).

2. In the current study, separate analyses for simple and complex structures were not conducted, to reduce complexity, as a significant number of factors were already examined.

Table 1. Verb category and conjugation

Category	Complexity	Uninflected form	Present progressive
AR	Simple	lavar ('laugh')	lav**iando**
ER	Simple	poner ('sleep')	pon**iendo**
IR	Simple	partir ('dance')	part**iondo**
AS	Complex	montas ('clean')	mant**iando**
ES	Complex	detenes ('read')	diten**iendo**
IS	Complex	recibis ('smoke')	rocib**iondo**

These uninflected verbs were real Spanish verbs to which meaning was arbitrarily assigned in order to prepare a sufficient number of action verbs. For instance, *lavar* means "to wash" in Spanish, but it means "to laugh" in *Supurango*.

Assessment tests

The assessment tests were ~~also~~ computerized and administered using the DMDX software (Forster & Forster, 2003), and the responses were audio-recorded. Two types of tests were used, namely (a) rule-application test and (b) PP test, without providing any feedback.[3] While the rule-application test was used to examine to what extent the participants learned the morphological rules independently from vocabulary knowledge, the PP test assessed the extent to which the participants could use the correct PP form of 24 verbs that they practiced. Participants' responses were scored by trained independent raters based on the scoring procedures used by Suzuki and DeKeyser (2017a). In training, two raters coded the same set of data (16% of each test) until their coding matched. In all two tests, accuracy of each response was scored as all or nothing. The order of the test items was randomized for each test at different time points (from Session 1 to Session 6), thus precluding the participants from having any expectations regarding the verb-type item that would be presented next.

Rule-application test

For the rule-application test, new verbs were created based on the verbs in the training sessions, replacing the phonemes of the stem but keeping the number of syllables (e.g., the practiced verb *lavar* was changed to nonce verbs such as *nopar*, for the list of nonce verbs). The task objective was to convert these unknown, uninflected verbs (e.g., *nopar*) to a PP form (e.g., *nopiando*) as quickly as possible. The participants heard a new uninflected verb through headphones and were

3. A vocabulary knowledge test was also conducted in Suzuki (2017); however, it was excluded from the current analysis because the focus of the current study was the acquisition of morphological structures.

shown the spelling on the screen. They were then asked to change it to the PP form within eight seconds, after which the next item was automatically presented. Twenty-four items (four verbs for each category) were created and used for the tests after the training (Tests 4B, 5 and 6). A different set of 12 verbs (two verbs for each category) were used for the tests during the training phase (Tests 1-4A). The number of the items was decreased during the training sessions to reduce the interference on the learning of the actual verbs (see the list of verbs in Appendix S2 in the online supporting information). It took approximately one minute and a half to complete the 12-item tests and three minutes to complete the 24-item test. Cronbach's alpha for all tests was acceptable, ranging from .85 to .94.

PP test
In the PP test, the participants were presented with still images in which a man was performing various activities. These images were snapshots taken from videos shown in the learning sessions, thus ensuring that the participants were familiar with their meaning. As in the rule-application test, eight seconds were given for each test item. It took approximately three minutes to complete the test. Cronbach's alpha associated with all tests was acceptable, ranging from .81 to .92.

Aptitude tests

Four aptitude tests were administered to assess three aptitude components, namely LAA, WMC, and RMA. LAA was measured by the LLAMA-F, a sub-test of the LLAMA aptitude test (Meara, 2005), while WMC was measured by the automated Ospan task adopted from Unsworth, Heitz, Schrock, and Engle (2005). On the other hand, RMA was assessed by the composite score of the two tests, the LLAMA-B (Meara, 2005) and LABJ-PA, a sub-test of the LABJ (the Language Aptitude Battery for Japanese) test (Sasaki, 1993).

LLAMA-F
In the LLAMA-F, the participants were required to infer the grammar rules by looking at pictures and word sequences. The test consisted of a learning phase and a test phase. In the learning phase, participants were given five minutes to learn a new language by examining sentences corresponding to different images. In the testing phase, they were given two sentences, one of which was grammatically correct and the other was not, and were instructed to choose the grammatical sentence. The test included 35 items. The original LLAMA-F test administered by Suzuki and DeKeyser (2017b) consisted of 20 test items and indicated a somewhat low reliability (Cronbach's alpha = .53). In the present study, fifteen items were added that were similar to several of the original items that showed relatively higher item-total correlations in the original sample. Cronbach's alpha in the present version was higher and was satisfactory ($\alpha = .75$).

Ospan tasks

In the Ospan task, for each item, participants first solved a math problem, indicating whether the solution of an equation was correct or incorrect. After each math problem, they were presented with a letter of the alphabet and asked to remember it. After each set of math problems and letters, they were asked to select the letters in the presented order. Successful performance on this task requires temporary updating of incoming information consecutively. The test was comprised of fifteen trials, with three trials for five sets with five different sizes each (3 to 7). The total number of sets was 75.

The Ospan task was scored as the sum of all correctly recalled letters in correct positions based on the procedure used in Unsworth et al. (2005). For example, if an individual correctly recalled four letters in a set size of six, the score was four. In order to make sure that the Ospan task was performed appropriately, only participants that solved the math problems with high accuracy rates are usually recommended for inclusion. The average accuracy rate was 90.76% ($SD = 6.00\%$, range: 69%–97%). In the study conducted by Unsworth et al. (2005), an 85% accuracy criterion (i.e., a maximum of 12 errors out of the 75 operations) was set for all participants. Six participants scored below the criterion (five in the shorter-spaced learning group and one in the longer-spaced learning group) and their scores were excluded from further analysis. Reliability indexed by Cronbach's alpha was acceptable ($\alpha = .68$).

LLAMA-B

In the LLAMA-B test, the participants were presented with 20 images (unique unfamiliar creatures) and were required to learn their names (i.e., real words from a Central-American language). They were required to remember as many words as possible during the two-minute study phase. In the test phase, they were presented with each of the 20 words and were asked to choose the correct image (creature). The total possible score was 20 and Cronbach's alpha was satisfactory ($\alpha = .77$).

LABJ-PA

In the LABJ-PA, the participants were required to remember a list of 24 Kurdish-Japanese word pairs. After the four-minute learning phase, they were tested via a multiple-choice test, whereby they were asked to select the correct meaning of each Kurdish word from the four options. The total possible score was 24 and Cronbach's alpha was satisfactory ($\alpha = .80$).

Research design and procedure

The study design was based on two experimental groups to which the participants were randomly assigned. Each participant took part in six sessions (four training sessions and two delayed test sessions) accompanied by the administration of the aptitude tests. They were randomly assigned to the following two groups that received identical treatments while their training intervals were manipulated, with 3.3-day interval given to the shorter-spaced learning group and 7-day interval for the longer-spaced learning group. As shown in Figure 1, participants in the shorter-spaced learning group engaged in the four training sessions either on Mondays and Thursdays or on Tuesdays and Fridays, whereas those in the longer-spaced learning group attended one session every week. During the first four training sessions, the participants completed the assessment tests, allowing their learning gains and losses to be tracked. The assessment tests (i.e., vocabulary, rule-application, and PP) were administered at the beginning of training Sessions 2–4 (Tests 2A, 3A, and 4A) for assessing forgetting and retention and at the end of each training session (Tests 1, 2B, 3B, and 4B) for assessing learning gains. In Sessions 5 and 6, no training was conducted, and the same set of tests was administered 7 days and 28 days after Session 4. These post-tests were given in order to examine the extent to which the aptitude effects persist after the training sessions. After the assessment tests were completed, the participants completed LLAMA-F and Ospan task as a part of Session 5, while LLAMA-B and LABJ-PA were given in Session 6.

Training procedure

All training sessions were controlled by the computer program (see Suzuki, 2017 for more detailed description of the procedure). Table 2 provides an overview of the training procedures. Following the procedure adopted by Suzuki and DeKeyser (2017a), the target structure was taught in an explicit step-by-step manner, comprising of (a) vocabulary learning, (b) understanding explicit grammatical explanations about the PP, and (c) oral practice using the PP.

In the vocabulary practice phase, the participants were presented with an image depicting an action verb (Japanese translation was shown in the top right corner as well) and were prompted to say the Supurango equivalent. Each image was presented on the screen for four seconds, followed by the feedback. Feedback was given in both oral and written forms, whereby the word was pronounced once while its written form remained on the screen for four seconds. The set of 24 verbs

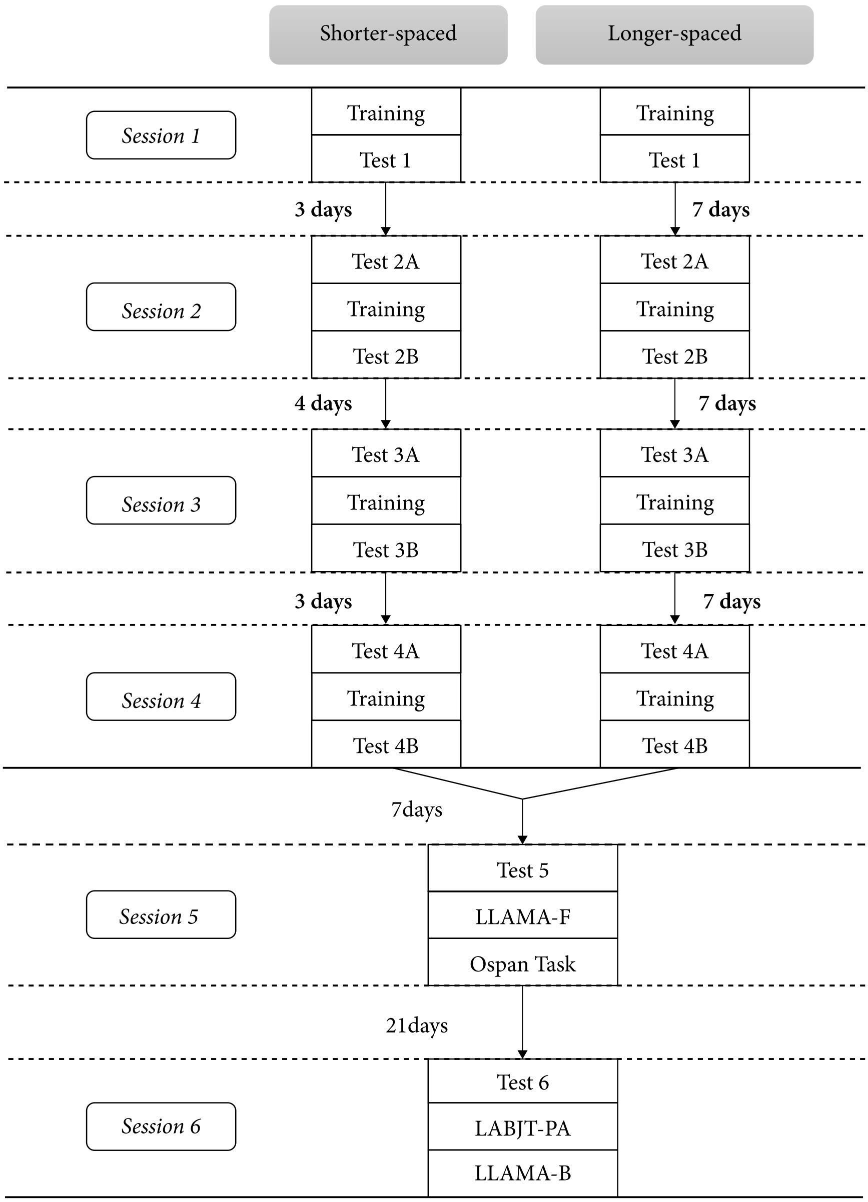


Figure 1. Research design
Note. Each Test session involved a rule-application test and a PP test, as well as a vocabulary test. Test A sessions occurred prior to the training session, whereas Test B sessions were conducted after the training session.

Table 2. Training procedures

Session 1		Session 2–4	
Task	Min.	Task	Min.
1. Questionnaire and Consent Form	5	1. Test A (2A, 3A, 4A)	7
2. Vocabulary Practice	14	2. Vocabulary Practice	16
3. Explicit Information Sheet and Explanation	5	3. Explicit Information Sheet	1
4. Grammar Practice	20	4. Grammar Practice	20
5. Test 1	7	5. Test B (2B, 3B, 4B)	7

was repeated three times in Training session 1 and four times in Training sessions 2 to 4.

After the vocabulary learning phase, the participants were provided with a sheet that explained verb conjugations (see Appendix S3 in the online supporting information). During the grammar practice in all the sessions, the participants were encouraged to refer to it as required. In Session 1 only, the participants also read a series of slides that explained the conjugations for each category after receiving the explicit information sheet.

In the grammar practice phase, the participants watched an animation video in which a man performed an action corresponding to the verbs they were required to master. Each video clip lasted eight seconds, and the participants had to orally describe the animation using the PP form of the verb. As in the vocabulary learning phase, feedback was given in both oral and written form.

Data analysis

Multiple regression analyses were conducted with aptitude test scores as predictors of the assessment test scores. Multiple regression analysis, rather than correlational analysis, was chosen as the primary analysis method in order to control for correlations among independent variables (Tabachnick & Fidell, 2007). The correlation results yielded are intended as a supplement to the results of multiple regression analyses (see Appendix S4). Two sets of 18 multiple regression analyses were conducted with the two assessment tests as dependent variables (9 rule-application and 9 PP tests) for longer-spaced and shorter-spaced learning groups, respectively. The model comprised of MRRA and WMC as predictors and generated multiple regression coefficients (unstandardized coefficient, *B*; standardized coefficients, β). No multicollinearity was detected in the data set, as the VIF was less than 10 and the tolerance was above .02 (Field, 2009).

With 18 multiple regression models being constructed, a guard against the likelihood of obtaining false-positive results (Type 1 errors) was needed. The significance of regression coefficients for the two predictors was tested using the Benjamini-Hochberg procedure with a 5% false discovery rate (Benjamini & Hochberg, 1995). The Benjamini-Hochberg procedure has more advantages than Bonferroni correction because the latter can often be too conservative (Bender & Lange, 2001). The results of the omnibus models are presented in the online supporting material (see Appendices S5 and S6). Based on the predictions generated by previous findings from Suzuki and DeKeyser (2017b), the correction procedure was applied to one of the predictors. Specifically, because MRRA was hypothesized to predict L2 learning only for the longer-spaced practice group, the correction was conducted on the 18 coefficients only of the 36 MRRA ones. Similarly, because WMC was hypothesized to predict L2 learning only for the shorter-spaced practice group, the correction was conducted on only half of the 36 WMC coefficients.

Results

Descriptive statistics

Assessment tests

Descriptive statistics for the three assessment tests are presented in Table 3. Results from inferential statistics are not reported here, as the aim of the current study was not ascertaining the difference between the two groups (see Suzuki, 2017 for the detailed results). The major pattern in the findings was that the shorter-spaced learning group outperformed the longer-spaced learning group consistently in the earlier stages of training. The advantage of the shorter-spaced learning condition was statistically significant ($p < .05$) in all test phases, from Test 3A to the final posttest, across all three assessments (vocabulary, rule, and PP).

Aptitude tests

Descriptive statistics and correlations among the aptitude test results are presented in Table 4. As can be seen, all aptitude test scores (Ospan task, LLAMA-F, LLAMA-B, and LABJ-PA) were significantly correlated to each other. The strongest, albeit moderate, correlation was found between LLAMA-B and LABJ-PA ($r = .56$, $p < .001$), both of which measured RMA. In the present study, a composite score of LLAMA-B and LABJ-PA (standardized z scores) was computed and used as an RMA score. The strength of the remaining correlation coefficients was weak ($.28 < r < .37$, $p < .05$). The second-order aptitude component (i.e., RMA

Table 3. Descriptive statistics of three tests for longer-spaced and shorter-spaced learning groups

	Shorter-spaced			Longer-spaced		
	N	*M*	*SD*	*N*	*M*	*SD*
Rule						
Test 1	30	48.89	30.46	30	36.67	30.84
Test 2A	30	20.00	30.37	30	11.94	21.74
Test 2B	30	69.17	26.73	30	59.72	31.11
Test 3A	30	63.61	31.98	30	36.94	34.09
Test 3B	30	85.83	16.69	30	65.00	32.49
Test 4A	30	79.17	22.61	30	51.67	32.93
Test 4B	30	87.92	13.73	30	72.50	31.00
Test 5	30	85.97	13.99	29	67.39	31.10
Test 6	30	79.31	15.22	30	65.00	28.96
PP						
Test 1	30	25.28	18.82	30	20.00	15.49
Test 2A	30	13.06	14.59	30	6.67	11.51
Test 2B	30	72.36	18.75	30	60.97	25.08
Test 3A	30	54.86	25.62	30	35.00	26.18
Test 3B	30	90.42	14.02	30	73.47	24.23
Test 4A	30	81.94	15.41	30	58.06	28.64
Test 4B	30	94.17	9.70	30	83.75	17.49
Test 5	30	85.69	14.83	29	70.40	23.68
Test 6	30	77.50	16.25	30	65.14	21.29

Note. The maximum possible score was 100.

and MRRA) also exhibited weak relations with the first-order aptitude components (Ospan and LLAMA-F). The Ospan score was positively related to RMA and MRRA ($r = .33$ and .41, $p = .014$ and .002), whereas LLAMA-F was also related to RMA ($r = .40$, $p = .002$).

In order to examine whether the aptitude scores of the two groups were comparable, independent-samples *t*-tests were conducted for six aptitude components. None of the comparisons produced statistically significant results at $p > .05$ (see Appendix S6 in the online supporting information for the results of *t*-tests).

Table 4. Means, standard deviations, and inter-correlations for aptitude test scores

	Ospan	LLAMA-F	LLAMA-B	LABJ-PA	RMA (LLAMA-B + LABJ-PA)	MRRA (RMA + LLAMA-F)
Ospan		.35**	.28*	.31*	.33*	.41**
LLAMA-F			.32*	.37**	.40**	.72**
LLAMA-B				.56**	.88**	.80**
LABJ-PA					.88**	.82**
RMA						.92**
MRRA						
M	55.72	27.67	12.72	18.83	.00	.00
SD	9.33	4.38	3.94	4.09	.88	.78

Note.
* $p<.05$ ** $p<.01$

Multiple regression analysis

Table 5 presents multiple regression coefficients of MRAA and WMC for the longer-spaced and shorter-spaced learning groups. In the longer-spaced learning group, MRRA predicted all the rule-test performance (with the exception of Tests 1 and 2B) and all PP test performance. WMC, in contrast, did not play any role in the test scores.

In the shorter-spaced learning group, neither MRRA nor WMC turned out to be a significant predictor of rule-application test performance. On the other hand, for the PP test performance, only MRRA showed a systematic pattern of associations. Specifically, MRRA was significant predictor of the scores on the PP tests, which were administered immediately after the participants completed the training sessions (i.e., Test 2B, 3B, and 4B).

Discussion

An overview of multiple regression results is summarized in Figure 2. The findings are presented and discussed for each of the two research questions.

Table 5. Multiple regression coefficients of MRRA and WMC: Models with MRRA and WMC as predictors[Shorter-spaced Learning Group]

	MRRA						WMC					
	B	*SE*	*β*	*t*	*p*	95% CI	*B*	*SE*	*β*	*t*	*p*	95% CI
Rule												
Test 1	−1.38	10.05	−0.03	−0.14	.89	[−22.22,19.47]	2.06	0.92	0.51	2.25	.03	[0.16,3.96]
Test 2A	4.38	10.91	0.10	0.40	.69	[−18.25,27.01]	1.01	0.99	0.25	1.02	.32	[−1.05,3.08]
Test 2B	17.99	8.74	0.46	2.06	.05	[−0.13,36.12]	0.26	0.80	0.07	0.32	.75	[−1.4,1.91]
Test 3A	−2.43	11.80	−0.05	−0.21	.84	[−26.9,22.03]	1.08	1.07	0.25	1.00	.33	[−1.15,3.31]
Test 3B	2.40	6.14	0.10	0.39	.70	[−10.34,15.14]	0.37	0.56	0.17	0.66	.51	[−0.79,1.53]
Test 4A	6.59	8.05	0.20	0.82	.42	[−10.09,23.28]	0.57	0.73	0.19	0.77	.45	[−0.95,2.09]
Test 4B	1.57	5.16	0.08	0.30	.76	[−9.14,12.28]	0.10	0.47	0.05	0.21	.84	[−0.88,1.07]
Test 5	5.57	4.57	0.27	1.22	.24	[−3.91,15.05]	0.56	0.42	0.30	1.34	.19	[−0.31,1.42]
Test 6	2.77	5.53	0.12	0.50	.62	[−8.7,14.24]	0.39	0.50	0.19	0.77	.45	[−0.66,1.44]
PP												
Test 1	10.05	6.59	0.37	1.53	.14	[−3.61,23.71]	0.08	0.60	0.03	0.13	.90	[−1.17,1.32]
Test 2A	−0.23	5.22	−0.01	−0.04	.96	[−11.07,10.6]	0.65	0.48	0.33	1.36	.19	[−0.34,1.63]
Test 2B	18.56	5.40	0.68	3.44	.00	[7.37,29.76]	−0.13	0.49	−0.05	−0.26	.80	[−1.15,0.89]
Test 3A	13.12	8.51	0.35	1.54	.14	[−4.54,30.77]	0.63	0.78	0.19	0.81	.42	[−0.98,2.24]
Test 3B	12.00	4.28	0.59	2.81	.01	[3.14,20.87]	0.02	0.39	0.01	0.06	.95	[−0.78,0.83]
Test 4A	9.39	5.17	0.42	1.82	.08	[−1.33,20.1]	0.16	0.47	0.08	0.34	.74	[−0.82,1.14]
Test 4B	8.68	3.06	0.61	2.83	.01	[2.33,15.03]	−0.15	0.28	−0.12	−0.55	.58	[−0.73,0.42]
Test 5	4.92	4.65	0.23	1.06	.30	[−4.71,14.56]	0.79	0.42	0.40	1.87	.08	[−0.09,1.67]
Test 6	10.34	5.36	0.44	1.93	.07	[−0.78,21.46]	0.19	0.49	0.09	0.40	.69	[−0.82,1.21]

Note. The shaded parts indicate that *p* was significant based on the Benjamini-Hochberg procedure with a 5% false discovery rate.

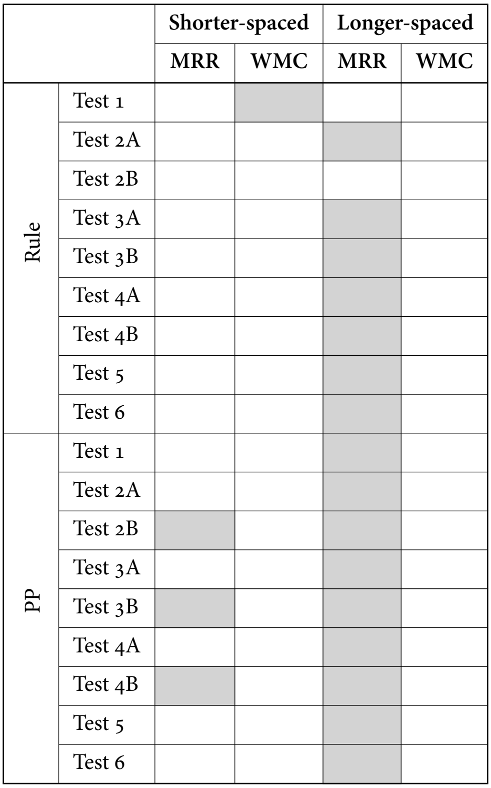
Table 5 *(continued)*

[Longer-spaced Learning Group]

	MRRA						WMC					
	B	*SE*	*β*	*t*	*p*	95% CI	*B*	*SE*	*β*	*t*	*p*	95% CI
Rule												
Test 1	6.25	7.29	0.17	0.86	.40	[−8.75,21.24]	0.17	0.60	0.06	0.29	.78	[−1.06,1.4]
Test 2A	12.38	4.64	0.48	2.67	.01	[2.85,21.91]	−0.42	0.38	−0.20	−1.11	.28	[−1.2,0.36]
Test 2B	13.22	6.92	0.36	1.91	.07	[−1,27.44]	0.22	0.57	0.07	0.39	.70	[−0.94,1.38]
Test 3A	18.32	7.39	0.46	2.48	.02	[3.13,33.51]	−0.30	0.60	−0.09	−0.50	.62	[−1.54,0.94]
Test 3B	18.34	6.95	0.48	2.64	.01	[4.06,32.61]	−0.61	0.57	−0.20	−1.08	.29	[−1.78,0.56]
Test 4A	20.79	6.81	0.54	3.05	.01	[6.78,34.8]	−0.52	0.56	−0.16	−0.93	.36	[−1.66,0.63]
Test 4B	19.85	6.32	0.54	3.14	.00	[6.87,32.84]	−0.82	0.52	−0.28	−1.60	.12	[−1.89,0.24]
Test 5	17.23	6.73	0.47	2.56	.02	[3.36,31.1]	−0.84	0.55	−0.28	−1.53	.14	[−1.98,0.29]
Test 6	16.26	6.15	0.48	2.64	.01	[3.62,28.9]	−0.74	0.50	−0.26	−1.46	.16	[−1.77,0.3]
PP												
Test 1	10.48	3.08	0.57	3.40	.00	[4.14,16.82]	−0.05	0.25	−0.03	−0.20	.84	[−0.57,0.47]
Test 2A	6.75	2.44	0.50	2.77	.01	[1.73,11.76]	−0.17	0.20	−0.15	−0.83	.41	[−0.58,0.24]
Test 2B	16.60	4.80	0.56	3.45	.00	[6.72,26.47]	0.30	0.39	0.13	0.77	.45	[−0.5,1.11]
Test 3A	17.19	5.29	0.56	3.25	.00	[6.31,28.06]	−0.11	0.43	−0.04	−0.25	.81	[−1,0.78]
Test 3B	17.23	4.49	0.60	3.83	.00	[7.99,26.47]	0.24	0.37	0.10	0.66	.51	[−0.51,1]
Test 4A	20.98	5.53	0.62	3.79	.00	[9.61,32.34]	−0.30	0.45	−0.11	−0.66	.51	[−1.23,0.63]
Test 4B	9.26	3.49	0.45	2.65	.01	[2.09,16.44]	0.39	0.29	0.23	1.38	.18	[−0.19,0.98]
Test 5	15.07	4.87	0.54	3.09	.00	[5.04,25.1]	0.07	0.40	0.03	0.18	.86	[−0.75,0.89]
Test 6	16.53	3.66	0.66	4.51	.00	[9,24.05]	0.25	0.30	0.12	0.82	.42	[−0.37,0.86]

Note. *p* is significant at the .05 level.

Figure 2. Summary of significant multiple regression coefficients



Note. Shaded cells indicate the significant multiple regression coefficients ($p < .05$).

Aptitude-treatment interaction patterns: Roles of MRRA and WMC

The first research question concerned the extent to which MRRA and WMC predicted L2 acquisition of morphological rules under the different practice-distribution conditions. The current study revealed that MRRA significantly predicted the rule and PP test scores consistently across time (with the exception for Tests 1 and 2B aimed at testing rule knowledge) in the longer-spaced learning group, not in the shorter-spaced learning group. The present findings lend support for the pattern reported by Suzuki and DeKeyser (2017b), who observed that LAA (a subcomponent of MRRA) was a significant predictor of the scores achieved only by the group provided a longer-spaced (7-day ISI) learning treatment. Learners with higher MRRA generally excel at instructed, explicit learning (Robinson, 2007) involving memorization of verbs and processing inflected verbs while rehearsing explicit grammatical rules. Therefore, they were most likely to have understood the morphological rules more deeply, which might have allowed them to retain the rules and inflected verbs even after a relatively longer-spaced (7-day) interval.

As opposed to the significant role of MRRA particularly on longer-spaced learning, the effect of WMC was not significant among the shorter-spaced learning group. The findings from Suzuki and DeKeyser's (2017b) study were not supported. Although the effects of WMC during the training phase were also examined, they were not significant as shown in test performance either in the beginning or at the end of each training session. Two possibilities are highlighted to seek future directions.

One possible explanation is simply that the meaningful association between WMC and learning under the shorter-spaced learning condition was spurious in Suzuki and DeKeyser's (2017b) research. The lack of association between WMC and L2 morphological learning is actually not inconsistent with the findings yielded by other empirical studies on morphological learning of L2 (artificial language). Although WMC was found to be associated with L2 morphology learning when no metalinguistic information is provided, WMC plays a limited role when grammatical explanations are given to study participants prior to morphology learning (Kempe & Brooks, 2008; Sanz et al., 2016). The effect of WMC was diminished in the current study, possibly because the metalinguistic information was provided during the training sessions. Hence, in future studies, it may be interesting to examine whether the role of WMC in longer-spaced and shorter-spaced learning conditions changes when metalinguistic information is not provided.

Another possibility may be more interesting and worthwhile to explore from the author's perspective. Recall that Suzuki and DeKeyser (2017b) interpreted the association between WMC and the learning outcomes attained by the shorter-spaced group as an indication of stronger memory interference experienced by the learners. These researchers claimed that similar morphological rules caused interference in their memorization, particularly in the shorter-spaced learning condition in which higher WMC helped to discriminate the similar morphological rules. Compared to the training interval employed by Suzuki and DeKeyser (1 day), the interval adopted in the current study (3.3 days) was longer. This longer interval might have lessened the interference and the burden on WM, yielding no systematic correlation between WMC and learning outcomes. The discrepancy between the research design and findings in these two studies raises an interesting question as to whether there is a threshold of intervals where the impact of aptitudes becomes substantial/insubstantial. In other words, a certain interval length that can alleviate the burden placed on individuals' WMC can potentially exist. Thus, authors of future research in this field should further examine the role of WMC in L2 learning by introducing a variety of training intervals, focusing on shorter intervals in particular.

Contribution of MRRA over the course of learning

The second research question pertained to the contribution of the aptitudes across time. MRRA was found to be a consistent predictor of all the stages of learning in the longer-spaced learning group. In contrast, MRRA played a more selective role in the learning gains exhibited by the shorter-spaced learning group (assessed by Tests 2B, 3B, and 4B). In other words, MRRA was *not* related to the scores achieved in the beginning of any of the training sessions (assessed by Tests 2A, 3A, and 4A) or after completion of the four training sessions (assessed by Tests 5 and 6).

Overall, the effects of aptitude did not greatly differ between the earlier and the later learning stages within this treatment period (RQ 2a; cf., Morgan-Short et al., 2014). In contrast, the effects of aptitudes were clearly selective when learning and forgetting/retention phases were distinguished (RQ 2b). Since MRRA pertains to the ability to remember and rehearse vocabulary and grammar rules for production, it is reasonable to posit that learning rate is predicted by MRRA in both groups. On the other hand, with respect to forgetting/retention of knowledge gained in the previous training session, MRRA significantly predicted the test performance in the longer-spaced learning group only, while it had no effect on the shorter-spaced learning group performance. Longer intervals (longer-spaced learning) seemed to have placed more burden on learners, whilst shorter intervals (shorter-spaced learning) seemed to have lessened the burden. It was shown that, among the longer-spaced learning group, learners with higher MRRA can retain newly acquired information in their memory better than those with lower MRRA. In contrast, shorter-spaced practice neutralized the effects of MRRA particularly on the forgetting/retention from the previous training session(s). In other words, shorter-spaced learning may be beneficial for L2 learners irrespective of their MRRA, i.e., leveling the playing field for different L2 learner types (Sanz et al., 2016). In sum, the findings yielded by the current study isolated the effects of aptitudes at different phases of L2 morphological learning.

Conclusion

The aim of the present study was to investigate the extent to which individual differences in cognitive aptitudes were associated with L2 morphological acquisition at different levels of practice distribution. Participants were trained on the morphological rules for oral production, and they completed four training sessions in either shorter-spaced learning (twice a week) or longer-spaced learning (once a week) schedules. The results yielded by the multiple regression analyses suggest that MRRA in particular has stronger predictive power. The role of WMC was found to be limited, contradicting the findings reported by Suzuki and DeKeyser (2017b). While the MRRA played an important role in all stages of learning in the longer-spaced learning group, the retention/forgetting of the previously trained items was not influenced by MRRA for the learners in the shorter-spaced learning group. This highlights that examining the roles of aptitudes during the training phase allowed for gleaning more fine-grained information about different stages of L2 learning processes.

These ATI findings suggest that practice distribution may be individualized based on learners' aptitude strengths to optimize L2 morphological learning (Vatz et al., 2013). For example, it can be inferred that shorter-spaced (shorter-interval) learning may neutralize the effects of individuals' MRRA on L2 morphological learning (Sanz et al., 2016). Recall that the group subjected to the shorter-spaced learning condition significantly outperformed participants assigned to the longer-spaced conditions, in Test 3A and all subsequent tests (see Suzuki, 2017 for details). Hence, when the aim is to maximize learning and retention, L2 learners can benefit most from a shorter-spaced learning schedule regardless of their MRRA.

The current study opens up new avenues for future research. First, in the work presented here, only two learning schedules (3.3-day and 7-day intervals) were employed. It would thus be beneficial to examine how other shorter (e.g., 1 or 2 days) and longer learning intervals (e.g., 14 days) influence the role of aptitude complexes. One possibility is that a longer interval monotonically leads to a greater effect of aptitude, whereas an alternative possibility is that the effects of aptitude plateau after a certain interval length. In order to establish these effects, authors of future research should employ more than two learning intervals and examine the function of learning intervals in aptitude complexes on a wider scale. Second, the current study was conducted as a laboratory experiment using an unfamiliar novel language structure. In order to seek higher ecological validity, the current findings need to be verified in more naturalistic learning settings (e.g., classroom instruction). Third, different types of aptitude complexes should be examined for learning of different grammatical structures (see Robinson, 2007 for different aptitude types). Since the role of aptitudes is moderated by target linguistic features such as saliency and difficulty (DeKeyser, 2012; Yalçın & Spada, 2016), in future research, authors need to select and examine different sets of aptitude complexes that are pertinent to the acquisition of target grammatical structures. Finally, future research in this domain can employ an off-line outcome task (e.g., a fill-in-the-blank task) to assess declarative grammatical knowledge, as well as procedural knowledge tasks, because optimal practice distribution may differ for the acquisition of declarative and procedural knowledge (Kim, Ritter, & Koubek, 2013; Suzuki & DeKeyser, 2017a; Ullman & Lovelett, 2018).

In sum, the current study revealed complex but systematic interactions between different levels of learning distribution and cognitive aptitudes. Its findings underscore that ATI researchers need to examine not only how aptitudes and treatment types interact but also establish when and why a certain ATI pattern is found, in order to uncover different L2 learning processes.

References

Anderson, J. R. (2015). *Cognitive psychology and its implications* (8th ed.). New York, NY: Worth Publishers.

Bender, R., & Lange, S. (2001). Adjusting for multiple testing – when and how? *Journal of Clinical Epidemiology*, 54, 343–349. https://doi.org/10.1016/S0895-4356(00)00314-0

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, 31, 635–650. https://doi.org/10.1017/S0142716410000172

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research*, 74, 245–248. https://doi.org/10.1080/00220671.1981.10885317

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). Rowley, MA: Newbury House.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning a temporal ridgeline of optimal retention. *Psychological Science*, 19, 1095–1102. https://doi.org/10.1111/j.1467-9280.2008.02209.x

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York, NY: Irvington.

DeKeyser, R. M. (2012). Interactions between individual differences, treatments, and structures in sla. *Language Learning*, 62, 189–200. https://doi.org/10.1111/j.1467-9922.2012.00712.x

DeKeyser, R. M., & Koeth, J. (2010). Cognitive aptitudes for second language learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 395–406). New York, NY: Routledge.

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. New York, NY: Routledge. https://doi.org/10.4324/9781315779553

~~Doughty, C. J. (in press). Cognitive language aptitude. *Language Learning, Early view*. https://doi.org/10.1111/lang.12322~~

Ellis, N. C., & Schmidt, R. (1997). Morphology and longer distance dependencies. *Studies in Second Language Acquisition*, 19, 145–171. https://doi.org/10.1017/S0272263197002027

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, 9, 147–172. https://doi.org/10.1191/1362168805lr161oa

Field, A. (2009). *Discovering statistics using SPPS*. London: Sage.

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, 34, 445–474. https://doi.org/10.1017/S0272263112000149

Granena, G. (2013). Cognitive aptitudes for L2 learning and the llama language aptitude test. In G. Granena, & Long, M. H. (Ed.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105–130). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.35.04gra

Granena, G. (2016). Cognitive aptitudes for implicit and explicit learning and information-processing styles: An individual differences study. *Applied Psycholinguistics*, 37, 577–600. https://doi.org/10.1017/S0142716415000120

Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language Learning*, 58, 703–746. https://doi.org/10.1111/j.1467-9922.2008.00477.x

Kim, J. W., Ritter, F. E., & Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science*, 14, 22–37. https://doi.org/10.1080/1464536X.2011.573008

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, 97, 634–654. https://doi.org/10.1111/j.1540-4781.2013.12030.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, 385–408. https://doi.org/10.1093/applin/amu054

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., … Doughty, C. J. (2013). Hi-lab: A new measure of aptitude for high-level language proficiency. *Language Learning*, 63, 530–566. https://doi.org/10.1111/lang.12011

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

Meara, P. M. (2005). *Llama language aptitude tests: The manual.* Swansea, UK: Lognostics. Retrieved from <http://www.lognostics.co.uk/tools/llama/llama_manual.pdf>

Miles, S. W. (2014). Spaced vs. massed distribution instruction for L2 grammar learning. *System*, 42, 412–428. https://doi.org/10.1016/j.system.2014.01.014

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong, P. C. M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, 17, 56–72. https://doi.org/10.1017/S1366728912000715

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition*, 37, 677–711. https://doi.org/10.1017/S0272263114000825

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult sla. *Language Learning*, 47, 45–99. https://doi.org/10.1111/0023-8333.21997002

Robinson, P. (2007). Aptitudes, abilities, contexts, and practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 256–286). New York, NY: Cambridge Universtiy Press. https://doi.org/10.1017/CBO9780511667275.015

Rogers, J. (2015). Learning second language syntax under massed and distributed conditions. *TESOL Quarterly*, 49, 857–866. https://doi.org/10.1002/tesq.252
Sanz, C., Lin, H.-J., Lado, B., Stafford, C. A., & Bowden, H. W. (2016). One size fits all? Learning conditions and working memory capacity in ab initio language development. *Applied Linguistics*, 37, 669–692. https://doi.org/10.1093/applin/amu058
Sasaki, M. (1993). Relationships among second language proficiency, foreign language aptitude, and intelligence: A structural equation modeling approach. *Language Learning*, 43, 313–344. https://doi.org/10.1111/j.1467-1770.1993.tb00617.x
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
Serrano, R., & Huang, H. Y. (2018). Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text? *TESOL Quarterly*, 52, 971–994. https://doi.org/10.1002/tesq.445
Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on esl learners' acquisition of articles. *TESOL Quarterly*, 41, 255–283. https://doi.org/10.1002/j.1545-7249.2007.tb00059.x
Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–94). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.2.06ske
Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 17–40). Amsterdam: John Benjamins. https://doi.org/10.1075/bpa.3.02ske
Snow, R. E. (1987). Aptitude complexes. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning and instruction* (pp. 13–59). Hillsdale, NJ: Lawrence Erlbaum Associates.
Suzuki, Y. (2017). The optimal distribution of practice for the acquisition of L2 morphology: A conceptual replication and extension. *Language Learning*, 67, 512–545. https://doi.org/10.1111/lang.12236
Suzuki, Y., & DeKeyser, R. M. (2017a). Effects of distributed practice on the proceduralization of morphology. *Language Teaching Research*, 21, 166–188. https://doi.org/10.1177/1362168815617334
Suzuki, Y., & DeKeyser, R. M. (2017b). Exploratory research on L2 practice distribution: An aptitude x treatment interaction. *Applied Psycholinguistics*, 38, 27–56. https://doi.org/10.1017/S0142716416000084
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). New York, NY: Harper & Row.
Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. *The Psychology of Learning & Motivation*, 60, 113–189. https://doi.org/10.1016/B978-0-12-800090-8.00004-4
Ullman, M. T. (2015). The declarative/procedural model: A neurobiologically-motivated theory of first and second language. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 135–158). New York, NY: Routledge.
Ullman, M. T., & Lovelett, J. T. (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research*, 34, 39–65. https://doi.org/10.1177/0267658316675195

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

Vatz, K., Tare, M., Jackson, S., & Doughty, C. (2013). Aptitude-treatment interaction studies in second language acquisition: Findings and methodology. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 273–292). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.35.11vat

Wesche, M. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 119–154). Rowley, MA: Newbury House.

Yalçın, Ş., & Spada, N. (2016). Language aptitude and grammatical difficulty. *Studies in Second Language Acquisition*, 38, 239–263. https://doi.org/10.1017/S0272263115000509

Yilmaz, Y. (2013). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics*, 34, 344–368. https://doi.org/10.1093/applin/ams044

Appendix

Online appendixes can be found at https://doi.org/10.1075/jsls.18023.suz.additional

Address for correspondence

Yuichi Suzuki
Faculty of Foreign Languages
Kanagawa University
3-27-1, Rokkakubashi, Kanagawa-ku
Yokohama-shi, Kanagawa, 221-8686
Japan
szky819@kanagawa-u.ac.jp

Author Query

- Please provide a complete reference for the citation '(Carpenter et al., 2012), (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006), (Sanz, Lin, Lado, Stafford, & Bowden, 2014), (Forster & Forster, 2003)' in this article.