Studies in Second Language Acquisition **0** (2017), 1–15
doi:10.1017/S0272263117000249

Research Report

THE ROLE OF PROCEDURAL LEARNING ABILITY IN AUTOMATIZATION OF L2 MORPHOLOGY UNDER DIFFERENT LEARNING SCHEDULES

AN EXPLORATORY STUDY

Yuichi Suzuki*

Kanagawa University

Abstract

This paper reports on the reanalysis of Suzuki's (2017) experiment and investigated the extent to which learning schedules influence automatization of second language (L2) morphology. Sixty participants were separated into two groups, which studied morphological rules for oral production under short-spacing (3.3-day intervals) and long-spacing learning conditions (7-day intervals). Their oral production test performance resulted in two measures of automatization: reaction time (RT) as an index of speedup and coefficient of variance (CV) as an index of stability/restructuring. The results showed that, while RT of both groups declined significantly after the training, the 3.3-day group exhibited greater propensity for restructuring than the 7-day group. Furthermore, procedural learning ability measured by the Tower of London task was significantly associated with RT, but not with CV, in the 3.3-day group only. These findings suggest that learning schedules and procedural learning ability influence different stages of automatization of L2 morphological learning.

INTRODUCTION

Attainment of automaticity is one of the goals of second language (L2) learning. Automaticity is defined as a fast, ballistic, effortless, and unconscious process (Segalowitz, 2003). Greater processing speed, indexed by shorter reaction time (RT) in a

This study was supported by Grant-in-Aid for Scientific Research (KAKENHI) from Japan Society for the Promotion of Science (JSPS). I am grateful to my RAs, Naoko Miyauchi, Yuki Aizawa, Misaki Kuratsubo, Yuka Oishi, and Aya Mizuno for their assistance in data collection and coding. I would like to express my gratitude to Kara Morgan-Short for allowing me to use the TOL task.

*Correspondence concerning this article should be addressed to Yuichi Suzuki, Faculty of Foreign Languages, Kanagawa University, 3-27-1, Rokkakubashi, Kanagawa-ku, Yokohama-shi, Kanagawa, 221-8686, Japan. E-mail: szky819@kanagawa-u.ac.jp

given task, is a hallmark of automaticity (DeKeyser, 2015). In addition to processing speed, Segalowitz and Segalowitz (1993) proposed stability as another dimension of automaticity. According to these authors, automatization entails some form of restructuring in language processing, leading to more stable processing. Unlike mere speedup of processing, restructuring entails elimination of inefficient processes and/or replacement of old, less efficient processes with new, more efficient processes. Restructuring/stability of language processing is indexed by the coefficient of variance (CV), which is computed by dividing the mean standard deviation (SD) of RT for each individual by his or her mean RT (Segalowitz & Segalowitz, 1993). The distinction between a mere speedup and restructuring has great potentials for L2 skill assessment (Hulstijn, Van Gelderen, & Schoonen, 2009).

The current study explores automatization of L2 morphological processing by examining the following three criteria: (a) faster RT, (b) smaller CV, and (c) a positive correlation between RT and CV. This assessment was guided by the premise that language processing becomes faster with practice, whereby the mean of both RT and SD declines. In other words, if the processing becomes "just faster," CV does not substantially change, as RT and SD decrease at a similar rate. Therefore, a combination of faster RT and unchanged CV can never produce a positive correlation. However, if the processing becomes more stable, it leads to a disproportional reduction in SD relative to RT, resulting in (b) smaller CV and (c) a positive correlation between RT and CV (see Segalowitz & Segalowitz, 1993 for detailed explanations). Meeting the criteria (b) and (c), in particular, indicates automatization in the narrow sense of restructuring, reflecting the ability to bypass or eliminate some of the processing components. These criteria have been applied for assessing automaticity in L2 grammatical processing (Hulstijn et al., 2009; Lim & Godfroid, 2015; Rodgers, 2011; Suzuki & Sunada, 2016). However, no research utilized these criteria to evaluate L2 automatization under different learning conditions in an intervention study. The current study probes L2 automatization under short-spacing and long-spacing learning conditions; it aims to contribute to a body of L2 research on distributed learning (Bird, 2010; Rogers, 2015; Suzuki, 2017; Suzuki & DeKeyser, 2017a).

In addition, this study explores the role of procedural memory for L2 automatization. Procedural memory system underlies cognitive and motor skill learning and involves gradual learning of "how," which contrasts declarative memory system for learning of "that," for example, remembering semantic and episodic knowledge (Squire, 2004). Procedural memory is involved in L2 learning, in particular for attaining more rapid and automatic grammatical processing (DeKeyser, 2015; Paradis, 2009; Ullman & Lovelett, 2016). Individuals' ability of learning within procedural memory (i.e., procedural learning ability) is often assessed by cognitive tasks such as a Tower of London (TOL) task (Kaller, Rahm, Köstering, & Unterrainer, 2011) and a serial RT task (Nissen & Bullemer, 1987).

A growing body of research has examined whether procedural learning ability plays a systematic role in L2 grammar learning (Antoniou, Ettlinger, & Wong, 2016; Brill-Schuetz & Morgan-Short, 2014; Ettlinger, Bradlow, & Wong, 2014; Hamrick, 2015; Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong, 2014). So far, all five empirical studies have yielded findings indicating that the outcome of training on (semi-)artificial language is related to procedural learning ability, suggesting that L2 adult learners recruit procedural memory for L2 grammar learning. In a study conducted by Morgan-Short et al. (2014), for instance, 14 native English speakers engaged in four training sessions on an artificial language (Brocanto2) over a two-week period. They took a grammaticality judgment test (GJT) in the middle and at the end of the training period. The study results showed that the participants' procedural learning ability, measured by the TOL task and the weather prediction task, predicted the GJT performance at the end of the training phase only. This pattern was replicated by another laboratory experiment performed by Hamrick (2015).

One of the limitations of the five studies noted in the preceding text is that their authors used accuracy scores on forced-choice receptive tests to assess L2 learners' grammatical knowledge, failing to measure another aspect of grammatical knowledge, that is, automatization or grammatical processing speed. Procedural memory acquires procedural linguistic knowledge gradually so that this knowledge will be processed rapidly and automatically in a later stage of acquisition (Ullman, 2015; Ullman & Lovelett, 2016). Therefore, speed measure may be able to tap into grammatical knowledge that is more conducive to automatization (i.e., gradual improvement in processing speed) than accuracy measure. In this context, it may be potentially even more interesting to examine whether mere speedup and restructuring may be distinguished by examining the relationships with procedural learning ability. However, none of the extant studies explored the role of procedural learning ability in RT (speedup) or CV (restructuring). The author of this study aimed to fill this gap in the extant knowledge by conducting an exploratory analysis to reveal the relationships of procedural learning ability with RT and CV.

In a related line of investigation, researchers have recently started to examine whether individual differences in procedural learning ability moderate the effects of different learning conditions (Brill-Schuetz & Morgan-Short, 2014; Tagarelli, Ruiz Hernandez, Moreno Vega, & Rebuschat, 2016). More specifically, Brill-Schuetz and Morgan-Short (2014) aimed to elucidate whether the role of procedural learning ability differed in explicit and implicit language training conditions. Participants in the implicit condition were simply exposed to artificial language (Broncanto2), while those in the explicit condition were taught pertinent grammatical rules before the exposure to the same language. Subsequent analyses of their performance revealed that procedural learning ability (a composite score of alternating serial reaction task and weather prediction task) predicted the posttest performance in the implicit condition only. In the more recent study conducted by Tagarelli et al. (2016), participants were exposed to a semiartificial language based on German syntactic rules under incidental or instructed learning conditions. The authors reported that, while procedural learning ability measured by the SRT task was not related to GJT performance in the instructed group, they noted a *negative* correlation between procedural learning score and GJT score. In sum, although these two studies (Brill-Schuetz & Morgan-Short and Tagarelli et al.) yielded contradictory findings, they do provide evidence indicating that the role of procedural learning ability may change depending on learning contexts, suggesting aptitude-by-treatment interaction patterns (Robinson, 2007). The aim of the current study was thus to explore the relationship between procedural learning ability and L2 grammar learning schedule.

THE CURRENT STUDY

The current paper reports on the reanalysis of Suzuki's (2017) experiment, the aim of which was elucidating the effects of learning schedule manipulation on L2 acquisition of morphology. In the present study, the author scrutinized automatization of L2 morphological processing under the short-spacing and long-spacing learning conditions. Sixty participants were trained on an element of a miniature language system (i.e., present-progressive morphological markers). Each participant took part in six sessions (four training sessions and two delayed test sessions). They were randomly assigned to the short-spacing or the long-spacing learning condition, whereby both groups received identical treatments while their training intervals were manipulated, 3.3-day and 7-day interval, respectively. As shown in Figure 1, participants in the 3.3-day interval group engaged in the four training sessions either on Mondays and Thursdays or on Tuesdays and Fridays, whereas those in the 7-day interval group attended one session per week. During the first four training sessions, the participants completed the assessment tests, allowing their automatization to be tracked during the training phase. The assessment tests were administered at the beginning of training sessions (Tests 2A, 3A, and 4A) and at the end of training session (Tests 2B, 3B, and 4B). In Sessions 5 and 6, the same set of tests was administered 7 days and 28 days after Session 4, respectively. The participants also completed the TOL task as a part of Session 6.

In Suzuki's (2017) experiment, results showed that the learners in the 3.3-day interval group were significantly more accurate in using the target structures than those in the 7-day interval group. However, no significant differences were noted between the two groups in terms of utterance speed (RT). Employing the theoretical framework of automatization (DeKeyser, 2015; Segalowitz, 2003), in this work, the original RT data yielded by the Suzuki's experiment is to further explore the automatization of L2 grammatical knowledge under the two different learning schedules. The current study is novel in two aspects, one of which pertains to the extension of the extant analysis of RT. More specifically, while individual's mean RT is used as a measure of *mere speedup*, CV, which is purported to assess *stability* or *restructuring* of processing, is also computed (Hulstijn et al., 2009; Lim & Godfroid, 2015; Rodgers, 2011; Segalowitz & Segalowitz, 1993; Suzuki & Sunada, 2016). The distinction between RT and CV is theoretically motivated, and is made in this work to advance our understanding of L2 automatization. Specifically, this study examined how automatization takes place during the training phase under two different learning conditions. While participants engaged in four training sessions, their performance was monitored by administering tests at the beginning and the end of each training session. Analysis focused on changes of RT and CV because of each of the training sessions.

Second, this study is also novel in that the author examined the role of procedural learning ability for L2 automatization under the short-spacing and long-spacing learning schedules. Prior research suggests that individual differences in cognitive aptitudes (e.g., language analytic ability and working memory) moderate the effects of different learning schedules on L2 morphological learning (Suzuki & DeKeyser, 2017b). By employing RT and CV as separate measures for speedup and restructuring, it is possible to explore the extent to which they are associated with individual differences in procedural learning ability. The aim of this investigation was to answer the following research questions:

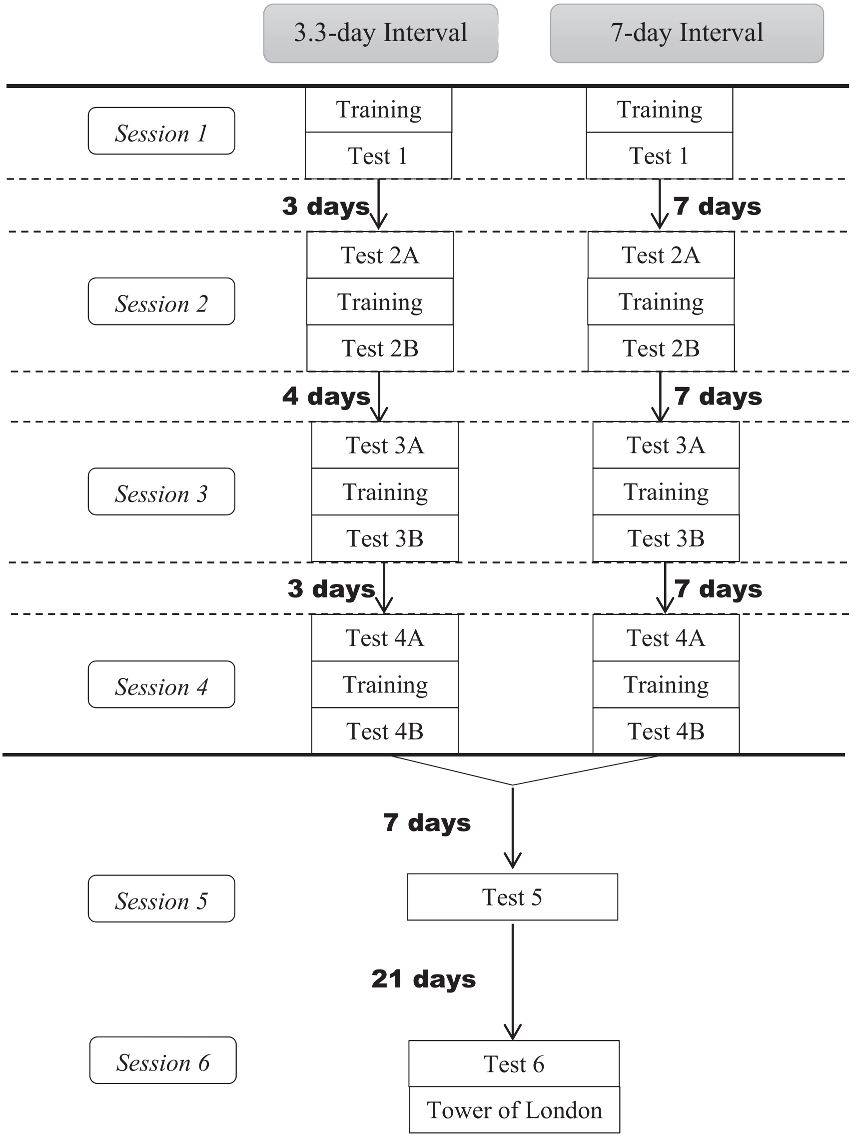

FIGURE 1. Research Design.

1. Do learners in the 3.3-day interval and the 7-day interval group show the evidence for automatization in the oral production tests (i.e., RT decrease, CV decrease, and a positive correlation between RT and CV)?
2. Is procedural learning ability related to RT and CV of the oral production tests?
3. Does the relationship among procedural learning ability, RT, and CV differ between the 3.3-day interval and the 7-day interval group?

METHODS

PARTICIPANTS

Sixty students attending a private Japanese university (20 of whom were male and 40 were female) participated in the study (see Appendix S1 for the participants' background information).

INSTRUMENTS

Target Structures

A miniature language called "Supurango" was created for the present study, which was loosely based on Spanish. Spanish was chosen because the words can be pronounced easily by Japanese speakers. None of the participants had any prior knowledge of Spanish language. The target grammatical structure was Present Progressive (PP), which was expressed by a morphological marking on a verb. As shown in Table 1, "Supurango" had six morphological rules depending on the verb ending. While three simple verb types (AR, ER, IR) required a change in a verb ending only, the other three complex verb types (AS, ES, IS) involved two transformations, that is, in the first vowel as well as in the verb ending.[1] Four action verbs were chosen for each verb type, producing 24 verbs for training.

Outcome Tests

The outcome tests were also computerized and administered using DMDX (Forster & Forster, 2003), and the participants' responses were audio-recorded. Three types of tests were conducted, namely (a) vocabulary test, (b) rule-application test, and (c) PP test. No feedback was provided to the participants. Because the focus of this study was the acquisition of L2 morphology, the vocabulary test results will not be reported.

Rule-Application Test

The aim of the rule-application test was assessing to what extent the participants learned the morphological rules independently from vocabulary knowledge. For this test, new verbs were created based on the verbs in the training sessions, replacing the phonemes of the stem while retaining the same number of syllables (e.g., the practiced verb *lavar* was

TABLE 1. Verb category and conjugation

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

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 saw the spelling on the screen, which remained on the screen for eight seconds. They were then asked to change it to the PP form. Twenty-four items (four verbs for each category) were created and used for the posttests. Different new verbs were created for each test; the items were different across the tests. Participants took approximately three minutes to complete the test.

Present Progressive Test

The PP test assessed the extent to which the participants could use the correct PP form of the same 24 verbs they practiced. They were presented with pictures in which a man was shown performing various activities. These pictures were still images adapted from the learning sessions, ensuring that the participants knew the meaning of each picture. As in the rule-application test, the prompt was presented for eight seconds for each item. The test took approximately three minutes to complete.

Tower of London Task

Procedural learning ability was assessed using a computer version of the TOL task (Kaller et al., 2011). The current TOL task was used as a significant predictor of the artificial language learning in the laboratory experiment by Morgan-Short et al. (2014). The same task procedure and scoring protocol as those employed in their experiment were adopted in the present study.

In this task, participants were presented with two configurations consisting of three pegs on which three colored balls were placed (see Appendix S2). The first and second configuration were labeled “Start” and “Goal,” respectively. Participants were instructed to move the balls on the pegs in a start configuration one at a time until it matches the goal configuration. They were required to achieve the goal configuration in the minimum number of moves, which was 3 to 6 depending on a particular trial. After completing a practice trial, participants solved a set of four 3-move trials, eight 4-move trials, eight 5-move trials, and eight 6-move trials. The duration of each trial was limited to one minute.

Three RT measures for procedural learning ability were obtained, namely (a) initial thinking time (ITT), (b) movement execution time (MET), and (c) overall solution time (OST). ITT refers to the time from the presentation of a trial to the first movement of a ball, and MET refers to the time from the first movement to the end of the last movement, while OST denotes the sum of ITT and MET. For these measures, the average proportional change scores were computed from the first to the final trial for each set of trials (i.e., 3-move, 4-move, 5-move, and 6-move trials). The formula employed for this calculation was as follows: (RT on the first trial − RT on the final trial)/RT on the first trial. Thus, the higher scores reflect a decrease in RT from the first trial to the final trial, implying greater procedural learning ability. Three proportional change scores were computed for ITT, MET, and OST.[2]

TRAINING PROCEDURE

All the training sessions were computerized and conducted using DMDX (Forster & Forster, 2003). Table 2 provides an overview of the training procedures. The participants engaged in deliberate practice of the target structures: (a) vocabulary learning, (b) understanding explicit grammatical explanations for the PP, and (c) oral practice using the PP (see Appendix S3 for details).

DATA CODING

For each test item, accuracy was scored as all or nothing (Suzuki & DeKeyser, 2017a). For the speed measures, RT was measured from the onset of the prompt to the end of the utterance. No RT measures were coded when (a) the response was incorrect and/or (b) the response included repairs and/or rephrasing. For a reliable computation of the average RT for each participant, the participants whose correct response rate (accuracy) was below 33% were excluded from the speed analysis (Suzuki & DeKeyser, 2017a). In addition, the outlying RT responses were excluded, following Suzuki and DeKeyser's (2017a) recommendation. More specifically, RTs below the minimum of 500 ms and RTs higher than 3 SD above the grand mean for each participant were treated as outliers, which accounted for 8.5% of the vocabulary tests, 1% for the rule-application tests, and 0.6% for PP tests. Internal consistency (reliability) indexed by Cronbach's alpha is provided in Appendix S4.

RESULTS

REACTION TIME AND COEFFICIENT OF VARIANCE DURING TRAINING PHASE

Accuracy scores are presented in Appendix S5 (see Suzuki, 2017 for detail). The learners' utterances during the tests were analyzed in terms of (a) reduction in RT, (b) decline in CV, and (c) a positive correlation between RT and CV. The changes in these indices from Test 3A to Test 3B and from Test 4A to Test 4B were examined. The analyses were not conducted on the change from Test 2A to 2B because the number of

TABLE 2. Training procedures

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

participants for whom such analyses could be performed was insufficient (8 and 4 learners in 3.3-day and 7-day groups for the rule test and 4 and 4 learners for the PP tests, respectively).

Table 3 presents the results for the rule tests. Paired *t*-tests[3] were conducted on RT and CV from Test 3A to Test 3B and from Test 4A to Test 4B for each group. To guard against Type I errors (false positives) from multiple t-tests, alpha levels were adjusted to .025 by using the Benjamini–Hochberg procedure with a 5% false discovery rate (Benjamini & Hochberg, 1995).[4] The results indicated that, except for Test 3A to Test 3B comparison in the 7-day interval group, the RT significantly decreased in both groups. CV declined significantly from Test 3A to Test 3B in the 3.3-day interval group only. There was a moderate positive RT-CV correlation on Test 3B in the 7-day interval group ($r = .43$, $p = .04$); however, this result does not reflect automatization (restructuring) because no changes in the RT were noted in Session 3. A weak positive correlation was found on Test 4B in the 3.3-day interval group ($r = .32$, $p = .09$).

Table 4 presents the results for the PP tests. Paired *t*-tests revealed significant RT reduction from Test 3A and 4A to Test 3B and 4B, respectively, in both groups. CV did not significantly change, except for a marginally significant difference noted from Test 4A to Test 4B in the 3.3-day interval group ($p = .08$). In both groups, weak positive correlations were found on Test 4B ($r = .34$, $p = .07$).

RELATIONSHIP OF PROCEDURAL LEARNING ABILITY WITH REACTION TIME AND COEFFICIENT OF VARIANCE

The current study was in part exploratory because I examined the relationships between three indices from the TOL task (i.e., ITT, MET, and OST) and test performance throughout the entire experiment (i.e., from Test 1 to Test 6). Given the number of correlation coefficients, some statistical correction procedure is usually required to lower the chances of Type I errors; however, no correction procedure was conducted because this part of study was exploratory (Bender & Lange, 2001). The magnitudes of correlation coefficients (r) were primarily interpreted based on the benchmark for L2 research (Plonsky & Oswald, 2014): small (.25), medium (.40), or large (.60). The significance values for p ($< .05$) should be interpreted only as preliminary evidence and a guide of the directions for future research.

Appendix S6 presents Pearson's correlation coefficients pertaining to procedural learning ability (ITT, MET, and OST) with RT and CV of the rule tests. Negative weak-moderate correlations were consistently found between MET and RT in the 3.3-day interval group across the tests ($-.33 < r < -.51$). On the Test 6, OST was significantly correlated with RT of the Test 6 ($r = -.38$, $p = .04$). None of the other correlations was meaningful or statistically significant.

Correlations of procedural learning ability with RT and CV of the PP tests are presented in Appendix S7. Consistent with the results obtained for the rule test, significant negative correlations were found between MET and RT only in the 3.3-day interval group. The ranges of the coefficients were similar to those related to the rule tests ($-.30 < r < -.51$). Unexpectedly, significant *positive* correlations, indicating association between higher procedural learning ability and greater CV, were found between ITT and CV of Test 5 in the 3.3-day interval group ($r = .42$, $p = .02$) and between OST and CV of

TABLE 3. RT and CV of rule test during training Sessions 3 and 4

		3.3-day ISI						7-day ISI					
		n	RT_Mean	RT_SD	CV	RT-CV correlation	*p*	n	RT_Mean	RT_SD	CV	RT-CV correlation	*p*
Test 3A		24	4125	1057	0.25	.30	.15	12	4318	1056	0.24	.06	.83
Test 3B		24	3576	733	0.20	−.06	.75	12	4301	942	0.23	.43	.04
t-test	*t*		5.55		2.46				0.10		0.43		
	df		23		23				11		10		
	p		.00*		.02*				.92		.68		
	Cohen's *d*		1.14		0.50				0.03		0.13		
Test 4A		29	4040	851	0.21	.20	.28	22	4203	964	0.23	−.03	.90
Test 4B		29	3584	847	0.23	.32	.09	22	3797	943	0.24	.18	.37
t-test	*t*		5.27		0.83				3.17		0.87		
	df		28		28				21		21		
	p		.00*		.42				.00*		.39		
	Cohen's *d*		1.06		−0.16				0.68		−0.19		

*The critical value was set to .025 for multiple comparisons (*t*-tests).

TABLE 4. RT and CV of PP test during training Sessions 3 and 4

		3.3-day ISI						7-day ISI					
		n	RT_Mean	RT_SD	CV	RT-CV correlation	*p*	n	RT_Mean	RT_SD	CV	RT-CV correlation	*p*
Test 3A		25	3835	1270	0.33	.06	.78	13	4032	1286	0.33	.30	.27
Test 3B		25	2865	890	0.30	.26	.16	13	3413	1137	0.34	.24	.22
t-test	*t*		6.76		1.37				2.61		0.44		
	df		24		24				12		12		
	p		.00*		.18				.02*		.67		
	Cohen's *d*		1.39		0.35				0.73		−0.23		
Test 4A		29	3662	1164	0.32	−.11	.57	25	3699	1197	0.32	−.06	.78
Test 4B		29	2695	781	0.28	.34	.07	25	2856	992	0.34	.34	.07
t-test	*t*		13.35		1.80				7.30		0.93		
	df		28		28				24		24		
	p		.00*		.08				.00*		.36		
	Cohen's *d*		2.53		0.36				1.59		−0.30		

*The critical value was set to .025 for multiple comparisons (*t*-tests).

Tests 4B and Test 5 in the 7-day interval group (r = .40, .39, p = .03, .04). Scatterplots for RT and MET (most of them showed significant correlations) are also presented in Appendices S8 and S9.

Given the exploratory nature of this study, I also present correlations of procedural learning ability with RT and CV of the vocabulary tests in Appendix S11 (for RT and CV changes of vocabulary tests during the training sessions, see Appendix S10) and those with accuracy scores on the three tests in Appendix S12. The interpretations of those results are beyond the scope of this study.

DISCUSSION

AUTOMATIZATION UNDER DIFFERENT LEARNING SCHEDULES

The current findings are summarized regarding the three criteria of automatization (reduction in RT, reduction in CV, and a positive correlation between RT and CV). First, in both groups, RT significantly decreased after the training. Second, descriptively, CV decreased in the 3.3-day interval group (except in the rule tests during Session 4), while virtually no reduction was noted in the 7-day interval group. Critically, leaners only in the 3.3-day interval group showed a significant reduction in CV on the rule test during Session 3. Third, a weak positive RT-CV correlation was found in both groups only on the PP test administered at the end of Session 4 (r = .34). In sum, although automatization during the training phase was not fully supported by the results related to either group, the data pertaining to the 3.3-day interval group provided greater evidence of automatization than that obtained for the 7-day interval group. This discrepancy indicates that more concentrated practice may facilitate automatization during the training, likely because more successful retrieval of the target items enabled learners to practice them more efficiently for automatization (see Suzuki & DeKeyser, 2017 for further discussion).

RELATIONSHIPS OF PROCEDURAL LEARNING ABILITY WITH REACTION TIME AND COEFFICIENT OF VARIANCE

The present findings suggest that procedural learning ability is associated with RT only, as no evidence of such association was noted for CV. Overall, the magnitudes of correlations were within weak to moderate range in terms of effect sizes across time, that is, from Test 2B to Test 6. Procedural learning ability seems to contribute to speedup, but not to more stable processing or restructuring (in the narrower sense of automatization). This result suggests that procedural learning ability plays a selective role in earlier stages of automatization (e.g., Segalowitz & Segalowitz, 1993). Later stages of automatization may not be influenced by procedural learning ability. Future research is thus needed to explore the role of other cognitive abilities, if any, that may account for individual differences in automatization beyond mere speedup.

It would be plausible to expect that the role of procedural learning ability would change depending on the amount of training, as some prior research suggests that procedural learning ability contributes to the acquisition of L2 grammatical knowledge assessed at the later stages of the training (Hamrick, 2015; Morgan-Short et al., 2014).

This assertion was partially supported in the present study. In the rule test (see Appendix S6), the magnitudes of correlation coefficients were the highest in Test 6 (r = - .51). However, this pattern was not observed in the results of the PP tests. Unlike the PP tests, the rule tests isolated the grammatical knowledge from vocabulary knowledge. Because procedural learning primarily supports grammar learning (Ullman, 2015), this study might have revealed a clearer association between procedural learning ability and grammatical knowledge, as assessed by the rule test.

In the present study, three measures pertaining to the TOL task (ITT, MET, and OST) were employed. However, only MET was consistently related to outcome measures.[5] Correlations pertaining to OST showed the same direction as those related to MET, but their magnitudes were weaker. In contrast, no correlation between ITT and the outcome was noted, which was inconsistent with the results of the experiment performed by Morgan-Short et al. (2014) who used the same TOL task. These authors found a meaningful association between GJT performance and ITT only, while noting no systematic relationship between GJT performance and MET or OST (Morgan-Short and Kate Brill-Schuetz, personal communication, February 19, 2016). Given these discrepancies, it can be postulated that MET might be a better measure of procedural learning. Unterrainer et al. (2004) found that, while better TOL performance was associated with *shorter* MET, it was related to *longer* planning time or ITT on difficult TOL trials (i.e., 7-move trials). Thus, improvements in ITT (i.e., shorter planning time) may not serve as a straightforward index of the procedural learning ability. In contrast, MET may involve motor learning ability, as well as online planning ability. Thus, the process that is tapped into by MET may be more compatible with the proceduralization of L2 learning because it involves both cognitive and motor skill learning (DeKeyser, 2015).

THE ROLE OF PROCEDURAL LEARNING ABILITY IN 3.3-DAY AND 7-DAY GROUPS

The systematic correlations between procedural learning ability and outcome test scores were found only in the 3.3-day interval group. These findings contribute to the extant body of knowledge on the acquisition of L2 grammar under different learning schedules, which is shown to recruit different types of cognitive abilities (Suzuki & DeKeyser, 2017b). CV analyses provided stronger evidence of automatization for the 3.3-day interval group relative to the 7-day group. Although a shorter-interval learning condition can be more advantageous for automatization, automatization might be more susceptible to individual differences in procedural learning ability in that condition.

Prior research indicated that procedural learning ability may be associated with learning outcomes assessed by GJT. However, this link pertains to the implicit (incidental) learning condition only, not the explicit (instructed) condition (Brill-Schuetz & Morgan-Short, 2014; Tagarelli et al., 2016). Because the training sessions employed in the present study involved provision of explicit grammatical rules, the findings yielded appear to contradict those reported in pertinent literature. When acquisition of grammatical knowledge is assessed through speed measures—rather than accuracy, as was the case in the studies conducted by Brill-Schuetz and Morgan-Short (2014) and Tagarelli et al. (2016)—procedural learning ability may play a significant role even among the explicitly instructed learners.

CONCLUSIONS

In the current study, automatization of L2 morphological processing was examined under the 3.3-day interval and 7-day interval learning conditions. The participants' automatization was assessed using RT (as an index of speedup in response) and CV (as an index of knowledge restructuring) obtained in the oral production tests. Although RT of both groups declined significantly after the training, the 3.3-day interval group exhibited slight advantages over the 7-day interval group in terms of CV. Furthermore, procedural learning ability significantly contributed to faster RT, but not to smaller CV. These significant associations were found in the 3.3-day interval group only. These findings suggest that learning schedules and individual differences in procedural learning ability influence different stages of automatization of L2 morphological learning.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S0272263117000249

NOTES

[1]Analysis of the data based on complexity of rules was beyond the scope of this study.

[2]One of the reviewers astutely pointed out that the current study could not calculate CV and used RT only to compute the change scores in the TOL task. This limited our analysis to the comparison between RT/CV from the language tests and RT from the TOL task. Fully crossed comparisons of RT and CV from both language and procedural memory tasks (i.e., RT-RT, RT-CV, CV-RT, and CV-CV) would have been possible if CV scores had been obtained by administering the TOL task twice to the same participants.

[3]Mixed ANOVAs with time (Session 3 and 4) and training (before and after completing the training tasks) as within-subject factor and with group as between-subject factor were conducted. because the number of participants was not equal across time and groups, the ANOVAs were performed only on the data pertaining to the participants whose responses were available across all the four tests (i.e., Test 3A, 3B, 4A, and 4B), resulting in very small numbers of cases. Here, the results of t-tests were mainly presented and interpreted.

[4]Benjamini-Hochberg procedure has more advantages than Bonferroni correction because the latter can be often too conservative (Bender & Lange, 2001).

[5]As pointed out by one of the reviewers, further research is needed to evaluate the robustness of the present preliminary findings using other measures of procedural learning ability (e.g., serial RT task).

REFERENCES

Antoniou, M., Ettlinger, M., & Wong, P. C. M. (2016). Complexity, training paradigm design, and the contribution of memory subsystems to grammar learning. *PLoS ONE*, *11*, e0158812.

Bender, R., & Lange, S. (2001). Adjusting for multiple testing—When and how? *Journal of Clinical Epidemiology*, *54*(4), 343–349.

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)*, *57*(1), 289–300.

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, *31*, 635–650.

Brill-Schuetz, K. A., & Morgan-Short, K. (2014). The role of procedural memory in adult second language acquisition. Paper presented at the 36th Annual Meeting of the Cognitive Science Society (COGSCI 2014), Quebec City, Canada.

DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 94–112). New York, NY: Routledge.

Ettlinger, M., Bradlow, A. R., & Wong, P. C. M. (2014). Variability in the learning of complex morphophonology. *Applied Psycholinguistics*, *35*, 807–831.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124.

Hamrick, P. (2015). Declarative and procedural memory abilities as individual differences in incidental language learning. *Learning and Individual Differences*, *44*, 9–15.

Hulstijn, J. H., Van Gelderen, A., & Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us? *Applied Psycholinguistics*, *30*, 555–582.

Kaller, C. P., Rahm, B., Köstering, L., & Unterrainer, J. M. (2011). Reviewing the impact of problem structure on planning: A software tool for analyzing tower tasks. *Behavioural Brain Research*, *216*, 1–8.

Lim, H., & Godfroid, A. (2015). Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, Van Gelderen, and Schoonen's 2009 study. *Applied Psycholinguistics*, *36*, 1247–1282.

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.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1–32.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Philadelphia, PA: John Benjamins Publishing Company.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912.

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 University Press.

Rodgers, D. M. (2011). The automatization of verbal morphology in instructed second language acquisition. *IRAL-International Review of Applied Linguistics in Language Teaching*, *49*, 295–319.

Rogers, J. (2015). Learning second language syntax under massed and distributed conditions. *TESOL Quarterly*, *49*, 857–866.

Segalowitz, N. S. (2003). Automaticity and second languages. In C. J. Doughty & H. M. Long (Eds.), *The handbook of second language acquisition* (pp. 382–408). Oxford, UK: Blackwell Publishers.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*, 369.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*, 171–177.

Suzuki, Y. (2017). The optimal distribution of practice for the acquisition of L2 morphology: A conceptual replication and extension. *Language Learning*. doi: 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.

Suzuki, Y., & DeKeyser, R. M. (2017b). Exploratory research on L2 practice distribution: An aptitude-by-treatment interaction. *Applied Psycholinguistics*, *38*, 27–56.

Suzuki, Y., & Sunada, M. (2016). Automatization in second language sentence processing: Relationship between elicited imitation and maze tasks. *Bilingualism: Language and Cognition*, 1–15. doi: 10.1017/S1366728916000857

Tagarelli, K., Ruiz Hernandez, S., Moreno Vega, J., & Rebuschat, P. E. (2016). Variability in second language learning: The roles of individual differences, learning conditions, and linguistic complexity. *Studies in Second Language Acquisition*, *38*, 293–316.

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. (2016). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research*. doi: 10.1177/0267658316675195

Unterrainer, J. M., Rahm, B., Kaller, C. P., Leonhart, R., Quiske, K., Hoppe-Seyler, K., et al. (2004). Planning abilities and the Tower of London: Is this task measuring a discrete cognitive function? *Journal of Clinical and Experimental Neuropsychology*, *26*, 846–856.