DOI: 10.1111/modl.12817

The Modern Language Journal

ORIGINAL ARTICLE

Effects of written languaging on second language learning: Mediating roles of aptitude

Masako Ishikawa[1] | **Wataru Suzuki**[2]

[1]Language Education Center, Josai University, Sakado-shi, Saitama, Japan

[2]Graduate School of Education, Miyagi University of Education, Aoba-ku, Sendai, Japan

Correspondence
Masako Ishikawa, Language Education Center, Josai University, Sakado-shi, Saitama, Japan 350-0295
E-mail: mi2185@josai.ac.jp

Wataru Suzuki, Graduate School of Education, Miyagi University of Education, 149, Aramaki-aza- Aoba, Aobaku, Sendai, Japan 980-0845
Email: suzukiw@staff.miyakyo-u.ac.jp

Languaging—that is, using language to mediate thinking, orally or in writing—facilitates second language (L2) learning. As languaging requires the deliberate, analytical action of L2 learners, learners with higher language aptitude are likely to benefit more from languaging than those with lower aptitude. Thus, this study investigated the extent to which language aptitude mediates the effects of written languaging (WL) on L2 learning. Pre-intermediate Japanese university students were assigned to +WL or −WL groups. The +WL group was instructed to compare their composition with the model text and explain, in writing, the use of articles (i.e., WL). In contrast, the −WL group copied the model 3 times. Examination of the association between the scores on aptitude tests (i.e., LLAMA, Language Aptitude Battery for Japanese, and Modern Language Aptitude Test) and the gain scores on assessment tests (pre-, post-, and delayed posttests) revealed that half of the correlations were significant for the 2 test types in the −WL group. However, no such correlations were found in the +WL group. Thus, WL might have benefitted the +WL participants by facilitating their noticing and reflecting on the linguistic issues, regardless of their aptitude profiles. We argue that WL may be a more influential factor than aptitude.

KEYWORDS
aptitude, noticing, reflection, remedial instruction, writing, written languaging

Like air, language is so natural that we overlook its significance, thinking of it as merely a tool for communication. However, language has more functions than that. People speak and write to clarify their thinking and remember things—that is, producing language helps people complete their thoughts and transform them into artifacts for further reflection, thereby mediating their cognition (Vygotsky,

1986, 1987). Informed by sociocultural theory of mind (SCT), Swain (2006) introduced the term "languaging" for these mediating functions of language and explained it as "the process of making meaning and shaping knowledge and experience through language" (p. 98). Pointing to the role of languaging as a learning tool, she further stated that "languaging about language is one of the ways we learn a second language to an advanced level" (p. 96).

Although the benefits of languaging apply to speaking and writing (Swain, 2006), written languaging (WL) has been underexplored compared to oral languaging (OL) despite its potential significance in learning. Comparing speaking and writing—"oral speech" and "written speech," respectively, in his terms—Vygotsky (1986) explained that they are separate linguistic processes "in both structure and mode of functioning" (p. 181). He further stated that writing encourages people to express their ideas more explicitly and elaborately, explaining, "Written speech is deployed to its fullest extent, more complete than oral speech (. . .) written speech requires what might be called deliberate semantics—deliberate structuring of the web of meaning" (p. 182). The differences may be attributable to two features exclusive to writing: slower pace and its visible product (Williams, 2012), which provide a valuable opportunity for people to see and reflect on their thinking. In the case of language learners, these features are likely to enable them to analyze their writing, mediating them to notice and reflect on any potential errors.

Languaging requires learners to solve their linguistic issues[1] by themselves, presumably using such intellectual resources as analytical abilities and grammatical sensitivity. Learners who are gifted with these abilities are thus hypothesized to benefit more from languaging than those who are not. If this speculation is correct, investigating the potential association would facilitate effective learning and teaching. Nonetheless, this link has rarely been examined (but see Ishikawa & Révész, 2023). In addition, although the positive impact of WL on learning has been reported (Suzuki, 2012), as mentioned before, WL is still underexplored compared to OL. Against this background, this study investigated the impact of WL on second language (L2) learning and the extent to which language aptitude moderates the effects of WL on L2 learning.

LANGUAGING

The concept of languaging originates from the perspective of Vygotsky's SCT, which views language as an essential mediator of cognition. According to Vygotsky (1987), "thought is not merely expressed in words; it comes into existence through them" (p. 219). Accordingly, sociocultural psychologists assume that humans develop through the internalization of language, which allows them to develop and regulate their thinking and behavior (Vygotsky, 1986, 1987; see also Lantolf & Poehner, 2023, this issue).

In L2 learning, supporting Vygotsky's argument that thinking is intimately related to language, Swain (2006) maintained that speaking and writing serve as "tools of the mind, mediating the cognition and re-cognition of experience and knowledge" (p. 106). In her view, producing language during tasks (i.e., languaging, both oral and written) leads to L2 development. Swain stated that the ways learners benefit from languaging are twofold. First, learners externalize their thoughts. Then, these externalized thoughts transform into artifacts for them to reflect on, enabling them to learn through the process and the product of their languaging. Put differently, languaging is a process that creates a visible and audible product about which one can language further (Swain, 2006).

WRITTEN LANGUAGING

As mentioned above, according to Swain (2006), the benefits of languaging accrue to both speaking and writing. Supporting her statement, second language acquisition (SLA) studies have produced evidence that both OL (Swain et al., 2009) and WL (Suzuki, 2012) mediate learning. Theoretically, WL is likely to have an equivalent mechanism to that of OL, in that learners externalize their thoughts

with language. That said, some differences arise from their respective modalities. According to Luria (1999)—an influential colleague of Vygotsky and the other key figure of SCT—unlike speaking, writing is a conscious process from the very beginning, as it takes special training to write. In addition, the following two characteristics are likely to facilitate learning: First, a slower pace of writing enables learners to focus solely on the act of writing at their own pace, usually under minimal time constraints (Williams, 2012). Therefore, linguistic processing, including noticing and metalinguistic reflection, may be "more likely to take place in writing than in speaking" (Manchón, 2011, p. 70). Second, the process of writing offers its content as a visible product without any additional recording device. Referring to both the slower pace and the products of writing, Williams (2012) stated that while some claims for the values of writing and output in general may overlap, "they may be stronger for written production due to the more generous time constraints and permanent record of writing" (p. 323). She further stated that learners are more likely to notice the holes or gaps in their production (Swain, 1998) in writing. In a similar vein, Luria (1999) contended,

> Written speech (. . .) assumes a much slower, repeated mediating process of analysis and synthesis, which makes it possible not only to develop the required thought but even to revert to its earlier stages, thus transforming the sequential chain of connections in a simultaneous, self-reviewing structure. (p. 103)

Supposing that the unique features of writing are likely to be carried over to WL, at least to some degree (see Gal'perin [1989], for communicated and dialogic thinking in his pedagogical model), the expected mediational role of languaging may be even more substantial in the case of WL. In addition, as a by-product of the slower pace, writing is likely to afford learners access to explicit knowledge (Williams, 2008), which "can be viewed as a 'tool' that learners use to mediate their performance and achieve self-control in linguistically demanding situations" (Ellis, 2009, p. 13). Given that explicit knowledge is considered conscious and potentially verbalizable (Ellis, 2004), it is likely to have significant importance for WL as it requires learners to verbalize their thoughts in writing.

In addition, as writing is "intrinsically an individual enterprise" (Manchón, 2011, p. 76), WL might enhance learners' potential to learn in the zone of proximal development (ZPD; Vygotsky, 1986) by facilitating the ability to develop independently. Although learners are expected to develop with the assistance of others in the ZPD (other-mediation), languaging is likely to function as a kind of mediation of the self (self-mediation). Namely, languaging moves what would normally be internal mediation (private speech) and externalizes it, making it more social and therefore more regulatory. The concept of using language as self-mediation has been shared by Gal'perin (1992), another contemporary of Vygotsky and major figure of SCT. Proposing his systematic theoretical instruction, Gal'perin aimed to expand learners' potential to learn in the ZPD by instructing them to verbalize instruction materials for them to shape "the ability to learn from something new independently" (p. 79). Given this, in the absence of "the expert other," learners might make use of WL, an artifact, to resolve their linguistic issues.

Finally, WL might have another benefit in the Asian context, at least in Japan. Probably because of "the teacher-centered nature of instruction" (Reichelt, 2009, p. 199), especially at the secondary level, and learners' shyness in general, as well as a decline in motivation after entrance examinations have been completed (Cummings, 2004), Japanese students are often unwilling to interact verbally (see Terauchi [2017] for an overview of English education at the tertiary level in Japan). Thus, they are likely to find writing less threatening than speaking (i.e., engaging in WL rather than in OL).

As stated, SLA studies have produced ample evidence that OL can facilitate learning (see Suzuki & Storch [2020], for a review). Starting with Suzuki's (2009) pioneering study, some studies have examined the possible positive impact of WL on learning (Suzuki, 2012, 2016). For example, in their partial replication of Swain et al.'s (2009) OL study, Ishikawa and Suzuki (2016) investigated the effects of WL on grammar learning using a pretest–posttest–delayed posttest design. Japanese English-as-a-foreign-language (EFL) learners were assigned to three groups: +WL, −WL, and control. After

treatment, the +WL group explained the grammar rules they learned, while the −WL group engaged in grammar exercises. Both the +WL and −WL groups outperformed the control group (which only took the pre- and posttests) on all posttests. Although no significant differences were identified between the two treatment groups, only the +WL group scored significantly higher on the delayed production posttest than the control group, suggesting a possible positive impact of WL on language development.

Similar results were obtained by Ishikawa and Révész (2020), who examined the effect of WL on the frequency and quality of WL with a pretest–posttest–delayed posttest design. Eighty-two Japanese EFL learners were divided into three groups: +WL, −WL, and control. All the groups took each test, which consisted of three assessments (i.e., essay test, grammar production test, and grammar recognition test) on the target construction, the English present counterfactual conditional (e.g., "If I could travel back to the past, I would dance with Michael Jackson"). The two treatment groups also participated in a 90-minute treatment session involving an individualized written dictogloss. After the learners reconstructed the text, the original text was distributed. The +WL group was instructed to compare the two texts carefully and write down their thoughts upon examining the text in Japanese (i.e., WL). In contrast, the −WL group was asked to compare the two silently (silent languaging) to avoid possible task effects that might have confounded the results of Ishikawa and Suzuki (2016). While it was certain that they did not write, they could well have engaged in OL (private speech) by whispering to themselves. The analyses of the gain scores of the three groups revealed that the +WL group significantly outperformed the control group, thereby excluding the possibility of test repetition effects. Comparison of the two treatment groups found that the +WL group scored significantly higher than the −WL group on the two production tests but not on the recognition tests. In addition, both the frequency and quality of WL were found to impact L2 development, with quality having a greater influence than frequency.

APTITUDE

Aptitude is assumed to be a specific talent for learning languages beyond the first, which exhibits considerable individual differences among learners (Dörnyei & Skehan, 2003). In addition, it has been considered to be a relatively fixed and stable trait (Carroll, 1981) and is strongly correlated with L2 proficiency (Ehrman & Oxford, 1995). As aptitude cannot be directly observed but has to be inferred from performance on psychological tests (Robinson, 2002), various aptitude tests have been developed, including the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959). Examining the data from the MLAT, Carroll (1981) identified four constituent abilities of aptitude: phonetic coding ability, grammatical sensitivity, rote learning ability, and inductive language-learning ability. However, inductive language-learning ability is not represented in any MLAT subtest.

Although aptitude information—that is, the results of aptitude tests—was initially used mainly for selection purposes, researchers eventually applied such information for other purposes, such as offering treatments that match learners' aptitude profiles (i.e., characteristics of learners' aptitude), to maximize the effectiveness of treatments, known as research on aptitude treatment interactions (ATI) in educational psychology. Cronbach (1967), an advocate of ATI, contended that instruction should be adapted to individual differences by altering the instructional method, suggesting that erasing individual differences is another way to adapt instruction to individual differences.

In the field of SLA, the significance of ATI has been increasingly recognized (see Li [2019] for a recent review). So far, SLA researchers have explored the interaction between aptitude and instruction, demonstrating that the types and nature of treatment (i.e., degree of explicitness or implicitness) can minimize differences in language aptitude (Li, 2013; Sheen, 2007). It should be noted, however, that previous studies have produced mixed findings. Some studies have reported fewer statistically significant correlations between aptitude and learning under explicit instruction than implicit instruction, suggesting that explicit instruction can minimize individual differences in aptitude (Li, 2013; Stefanou & Révész, 2015). Other studies, however, have shown more significant correlations between

aptitude abilities and development under explicit instruction, indicating that learners need both higher aptitude and explicit instruction to improve their development (Sheen, 2007; Yilmaz, 2013).

Regarding this inconsistency, Li (2013) explained that the difficulty of linguistic targets constrains the association between aptitude and the impact of feedback. In his view, when "the linguistic target is within learners' processing capacity" (p. 647) (i.e., simpler), they are likely to resolve their linguistic issues with explicit feedback regardless of aptitude, resulting in fewer significant correlations between explicit feedback and aptitude. In contrast, when the target is outside their processing capacity (i.e., more difficult), they are likely to need both explicit feedback and higher aptitude, resulting in significant correlations between the two (see Skehan [2015] for a similar discussion, and Yalçin & Spada [2016] for supporting evidence).

It should be pointed out, however, that the two studies by Sheen (2007) and Stefanou and Révész (2015) employed the same target (i.e., English articles) and produced inconsistent findings. Stefanou and Révész identified a significant correlation between aptitude and implicit—but not explicit—feedback in Greek high school students who were expected to have some knowledge of the target language feature. Meanwhile, Sheen found significant correlations between aptitude and explicit feedback in ESL learners who were not explicitly taught the target feature, indicating that they needed explicit feedback and higher aptitude abilities to improve. Given this, learners' prior knowledge of the target construction is likely to be another key.

Moreover, task types have been shown to influence interactions between aptitude and treatment. For example, Li et al. (2019) found that, regardless of their aptitude, participants who received explicit instruction (e.g., pretask instruction and within-task feedback) improved their scores on a grammaticality judgment test (GJT), a test of explicit knowledge, but not on an elicited imitation test (EIT), a test of implicit knowledge. The findings suggest that explicit instruction neutralized differences in aptitude on the GJT but not EIT, indicating task effects on learning.

Given these findings, it is important to elucidate the link between the impact of WL and aptitude to facilitate effective learning and teaching. As WL requires learners to resolve their linguistic issues independently, potentially functioning as self-mediation, learners with higher aptitude abilities, such as grammatical sensitivity and inductive language-learning ability, were hypothesized to benefit from WL more than those with lower aptitude. Nonetheless, to the best of our knowledge, Ishikawa and Révész (2023) is the only study to examine the possible link. This study examined the association between two types of grammar tests (a production test and a recognition test) and three aptitude tests (i.e., the MLAT, Language Aptitude Battery for Japanese [LABJ], LLAMA), which were also used in the current study. The analyses of the gain scores on the pre- and posttests and aptitude test scores of 64 participants assigned to the +WL group and −WL group revealed more significant correlations between the pre- and posttest gains and aptitude test results for the −WL group than the +WL group. Although rather unexpected, these results indicate that, in general, language aptitude played a stronger facilitative role for the −WL group than the +WL group. The researchers thus stated that WL might have leveled out the differences in aptitude, as it is likely to have enhanced the participants' noticing of the target form, facilitating their explicit learning—that is, learning with awareness (Leow, 2015)—and enabling them to draw on their explicit knowledge, resulting in improvement regardless of their aptitude.

As stated, target difficulty has been identified as a mediating factor. Against this background, it was deemed essential to examine the observed association between language aptitude and development from WL with target structures besides counterfactual conditionals, the focus of the earlier study by Ishikawa and Révész (2020). Accordingly, this study addresses the following research questions with articles as the target feature:

RQ1. To what extent does WL facilitate language learning?
RQ2. To what extent does language aptitude moderate the effect of WL on L2 learning?

TABLE 1 Background information per group

Group	*N*	Age		Gender		Years of English Study	
		M	*SD*	Male	Female	*M*	*SD*
+WL	18	18.6	0.5	10	8	7.2	1.8
−WL	17	18.5	0.5	9	8	7.5	1.7

Abbreviations: +WL, written languaging group; −WL, no written languaging group.

METHOD

Participants

The participants were 73 Japanese EFL learners at a private university in Japan. They were enrolled in one of three required first-year English classes, and the first author was their instructor. Two of the classes were the third and fifth highest levels of the 12 English classes in the management department. The other was the seventh highest level of the nine English classes in the pharmacy department. The participants were placed in these classes based on their scores on an in-house test administered at the beginning of the school year. There is a gap in English proficiency between the two departments, where pharmacy majors generally show higher proficiency than management majors. Before the study, the participants had studied English for 7.3 years on average, including 6 years of traditional grammar-oriented education in high school. Based on the results of the pretest and the aptitude tests, the participants were assigned to the +WL or −WL group through stratified random sampling. The initial number of participants was 73, but 38 participants who scored above the cutoff of 80% on the pretests and/or missed one of the posttests were eliminated. The final number of participants was 35. Their background information is summarized in Table 1.

Design

As summarized in Figure 1, the study was conducted with a pretest–posttest–delayed posttest design over 5 weeks during participants' regular class time. The participants assigned to the +WL and −WL groups took each of the language tests and the aptitude tests. In Week 1, informed consent and background questionnaires were administered. In addition, 15 vocabulary items that were likely to be unfamiliar to the participants were pretaught to ensure that they could focus on grammar without being confused because of the semantic difficulty of the items. In Week 2, students completed the aptitude tests. To experiment efficiently and effectively, the first author demonstrated with an overhead projector how to address the tasks with sample items that did not include the target construction. In Week 3, the pretest was administered. In Week 4, the treatment was conducted, followed by a posttest. Finally, the participants took the delayed posttest in Week 5.

Linguistic target

This study targeted two functional uses of the English article system: the indefinite article "a" for referring to something for the first time (first mention) and the definite article "the" to reference something already mentioned (anaphoric mention; e.g., "I saw a good movie yesterday, and I will recommend the movie to my friends"). Although it is less syntactically complex than the present counterfactual conditional employed by Ishikawa and Révész (2020), articles are considered challenging for learners because of their nonsaliency and low communication value (Celce-Murcia & Larsen–Freeman, 1999). In addition, they are especially difficult for learners whose first languages (L1s), including

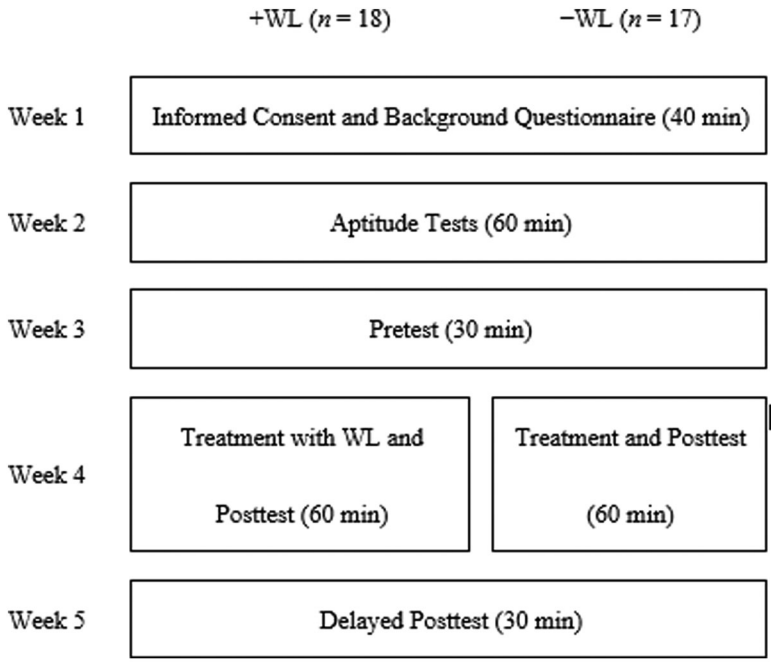

FIGURE 1 Flow of the procedure for the two groups.
Abbreviations: +WL, written languaging group; −WL, no written languaging group.

Japanese, do not have equivalent article systems (Master, 1997). As shown in Example 1, unlike English, Japanese does not require articles.

EXAMPLE 1

Watashi wa kuruma wo motte imasu
I car have
'I have a car.'

Thus, probably because of the L1–L2 difference, instances of inaccurate understanding and use of articles (e.g., overuse in nonobligatory contexts, omission in obligatory contexts) were observed in the first author's classes. The participants, however, had a little over 7 years of English education on average, and "a"–"the" distinction of the target was supposed to be familiar to them. Therefore, it was deemed appropriate for this study, where the participants were expected to learn the grammar rule for articles by themselves in only one 5-minute treatment.[2]

Assessment tasks and scoring

A production test and a recognition test were administered at each testing session to assess learners' knowledge of the meaning-to-form and form-to-meaning mapping associated with the target construction. Based on Muranoi's (2000) four-picture oral and written description tests, both tests were created with five pictures. For each test type, three comparable versions that differed only in pictures were developed (see Online Supporting Information A). Seven learners, similar to the participants in the current study in terms of proficiency and background, completed all the versions in counterbalanced order. A series of repeated-measures ANOVAs of their results showed no statistically significant differences among the test versions, production test, $F(2, 12) = 0.953$, $p = .413$, partial

$\eta^2 = 0.137$; recognition test, $F(2, 12) = 0.161$, $p = .853$, partial $\eta^2 = 0.026$. Three sets of tests were administered in a split-block design with a counterbalanced order. The participants were instructed to work on the production test first, followed by the recognition test for 12 minutes in each testing session. Dictionary use was not permitted.

Each version of the production test contained five pictures that provided participants with contexts to use the appropriate articles. The participants were asked to describe each picture using the words provided. The tests were coded for suppliance in obligatory contexts (Pica, 1983). First, the number of correctly used articles was counted and then divided by the number of obligatory contexts, yielding a percentage accuracy score. The first author coded all production tests twice in 6 months. The Cohen's kappa values for identifying the number of correctly used articles and obligatory contexts were 0.91 and 0.87, respectively.

Each version of the recognition test contained five pictures, below which sentences were provided with 20 multiple choices for articles (i.e., *a*, *an*, *the*, ∅) to describe each picture. The participants were instructed to choose the most appropriate option to complete each sentence. One point was awarded for each correct response, with 20 points being the maximum score. When "a" was selected in a context requiring "an," a partial point (0.5) was given.

Aptitude measures

As mentioned, three aptitude tests were employed to measure language aptitude: an adapted version of the MLAT Part IV (for grammatical sensitivity), the LLAMA_F, and the LABJ Part 2 (for inductive language-learning ability). All the tests were conducted in one of the computer rooms at the school. Below, we briefly describe the three tests according to the order in which they were conducted.

The MLAT Part IV measures the test-taker's awareness of the syntactic patterning of sentences and the grammatical functions of individual elements in a given sentence (Carroll, 1981). The MLAT was initially developed for test-takers whose L1 is English. Following this, the adapted version was created with the participants' L1, Japanese, to assess their grammatical sensitivity independent of their English proficiency (see Online Supporting Information B). Following the original format, each item consisted of two sentences. The first sentence included an underlined word or phrase, followed by a second sentence in which five words or phrases were underlined. The participants were instructed to identify the word or phrase in the second sentence with the same grammatical role as the underlined word in the first sentence. One point was given for each correct answer, and the maximum possible score was 20. No time limit was set, but most participants finished it in 3 minutes (range: 2–5 minutes). The participants who completed the test early were instructed to read their course textbooks. Hereafter, MLAT Part IV is referred to as "MLAT."

The LABJ (Sasaki, 1996) is a three-part test instrument developed specifically for Japanese speakers based on the MLAT (Part 1 and Part 3) and the Pimsleur Language Aptitude Battery (PLAB; Pimsleur, 1966; Part 2). As stated, inductive language-learning ability—that is, one of the four constructs identified by Carroll (1981)—was not included in any subtests of the MLAT. Thus, Sasaki (1996) used the translation of PLAB Part 4, which taps into inductive language-learning ability, as LABJ Part 2. It consists of 15 multiple-choice questions, and only this part was used in this study (for copyright reasons, no sample of the LABJ can be included in this article). So far, several SLA studies have used this test with Japanese university students similar to the current study's participants (Robinson, 2005; Shintani & Ellis, 2015). First, test-takers are instructed to examine examples of an artificial language written in the Japanese alphabet *katakana* and learn about that language for 1 minute. Then, based on the knowledge they acquire, they are instructed to infer the grammatical rules of the artificial language based on a set of words and sentences presented with their Japanese translations. A Japanese sentence is presented in each question, and participants are asked to choose the correct translation into the artificial language from the four choices provided. One point was awarded for each correct answer,

yielding a maximum score of 15. The time limit was 6 minutes, and the entire procedure took about 15 minutes. LABJ Part 2 is hereafter denoted "LABJ."

The LLAMA language aptitude test (Meara, 2005) is a computer-based test modeled on the MLAT. Unlike the original test, an artificial language is used that is accessible to people with any L1. The LLAMA consists of four subtests, but only LLAMA_F, a test of grammatical inferencing, was used. Similar to the LABJ, the subtest consists of two phases. In the first phase, test-takers are instructed to learn about the artificial language in 5 minutes. The second phase has 20 test items, each displaying a picture and two sentences describing the picture. One is correct, whereas the other is not. Thus, test-takers must choose the one based on the grammar rules learned in the first phase. When test-takers finish the test, their scores, ranging from 0 to 100, are displayed on the computer screen. Although the second phase is not timed, most participants finished the test in 6 minutes on average (range: 3–10 minutes). Again, participants who finished the test early were instructed to read their course textbooks. Henceforth, LLAMA_F is referred to as "LLAMA."

Treatment task and procedure

The treatment employed a picture description task with the same format as the production test. Based on Sheen's (2007) study, Aesop's fable about the crow and the fox was employed (see Online Supporting Information C). The rationale for using picture description was that it enabled us to prepare the text with the target construction. In addition, it ensured the participants' use and exposure to the target. The procedure consisted of three phases. First, a 3-minute Japanese video was shown to ensure that the participants knew the storyline. Second, a task sheet was distributed to the participants. They were instructed to describe, in writing, four pictures with the words provided next to each picture in the same manner as the production test. The time limit was 8 minutes. Third, a sheet with a model text was distributed to the participants, who were asked to carefully compare it with their descriptions, paying particular attention to the use of articles.

The sheet differed slightly depending on the treatment condition. For the +WL participants, this included the model text with 13 underlined articles and nouns. The participants were asked to explain, in writing, why each article was used or not used for each situation in Japanese (i.e., WL; see Online Supporting Information D). Meanwhile, the sheet for the −WL participants instructed them to copy the original text three times, thinking about why each article was used in the sentences.[3] The copying task was employed because copying involves writing, creating a condition similar to that of the +WL group.[4] Five minutes were allotted for this phase.

Statistical analyses

The data were analyzed with SPSS 26.0, with the level of significance set at 0.05.

First, descriptive statistics for the pre-and posttests of the two assessments (i.e., production and recognition tests) and three aptitude tests (i.e., MLAT, LABJ, and LLAMA) were calculated. As one of the distributions diverged from normality, nonparametric inferential statistical analyses were conducted to interpret the data. Accordingly, medians and interquartile ranges were employed as measures of central tendency and variation, respectively.

To investigate RQ1, the pretest scores were compared using a series of Mann–Whitney tests to confirm no significant differences between the groups at the outset of the study. Then, another series of Mann–Whitney tests was performed to assess any significant differences in the pretest–posttest and pretest–delayed posttest gain scores of the groups on the two assessments.

For RQ2, a series of Spearman's correlation analyses was conducted to identify the relationships between the participants' scores on the three aptitude tests and gain scores between the pretest–posttest

TABLE 2 Descriptive statistics for the two assessments by group and testing session

	+WL Group (n = 18)			−WL Group (n = 17)		
Test	***Mdn***	**IQR**	**95% CI**	***Mdn***	**IQR**	**95% CI**
Production test						
Pretest	0.11	0.25	[0.00, 0.24]	0.08	0.28	[0.00, 0.22]
Posttest	0.67	0.54	[0.39, 0.78]	0.28	0.50	[0.22, 0.61]
Delayed posttest	0.47	0.51	[0.25, 0.72]	0.10	0.39	[0.05, 0.39]
Recognition test						
Pretest	11.00	5.00	[8.00, 13.00]	10.00	7.00	[7.00, 14.00]
Posttest	16.50	7.00	[12.00, 17.75]	11.00	11.00	[7.00, 18.00]
Delayed posttest	15.00	9.00	[11.00, 19.00]	12.00	10.00	[8.00, 18.00]

Abbreviations: CI, confidence interval; IQR, interquartile range; *Mdn*, median; +WL, written languaging; −WL, no written languaging.
Note. Maximum score for production test was 1.0 (i.e., 100% accuracy); maximum score for recognition test was 20 points.

and pretest–delayed posttest on the two assessments. The effect sizes for the analyses were determined following Plonsky and Oswald (2014): d values of 0.40, 0.70, and 1.00 and correlation coefficients of 0.25, 0.40, and 0.60 were interpreted as small, medium, and large, respectively.

RESULTS

Pre- and posttest results

Table 2 presents the descriptive statistics for the pretest, posttest, and delayed posttest scores of the two assessments (i.e., production and recognition tests). On the production test, both groups scored around 0.1 point per context on the pretest, increasing their posttest scores and decreasing them on the delayed posttest. However, it is worth mentioning that even though both groups showed the same trajectories, the +WL group scored slightly higher on the pretest, and the gap between the two groups widened at the two posttests (as reported later, no significant differences were observed at the outset of the experiment). Namely, the posttest score of the +WL group increased to 0.67, whereas that of the −WL group was 0.28. Similarly, on the delayed posttest, the +WL group scored almost five times higher than the −WL group despite the decrease from the posttest.

On the recognition test, the +WL group scored 11.0 on the pretest, then raising their score considerably on the posttest at 16.5 but showing a slight decrease on the delayed posttest. Meanwhile, the −WL group scored 10.0 on the pretest, increasing their scores slightly on the two posttests.

A series of Mann–Whitney tests was performed to identify any differences between the groups at the outset of the experiment. The results demonstrated no significant differences in the production, $z = -0.390$, $p = .708$, $d = 0.132$; or recognition tests, $z = -0.066$, $p = .961$, $d = 0.022$.

Then, to address RQ1, another series of Mann–Whitney tests was run to assess whether the groups showed any difference in their pretest–posttest and pretest–delayed posttest gains on the two assessments by comparing the gain scores of the two groups. For the production test, a significant difference with a large effect size was identified for the pretest–delayed posttest gain score, $z = -2.809$, $p = .004$, $d = 1.077$; but not for the pretest–posttest gain score, $z = -1.350$, $p = .184$, $d = 0.471$. Meanwhile, for the recognition test, a medium-sized significant difference was observed for the pretest–posttest gain score, $z = -2.134$, $p = .032$, $d = 0.774$; but not for the pretest–delayed posttest gain score, $z = -1.541$, $p = .126$, $d = 0.539$.

TABLE 3 Descriptive statistics for the aptitude tests

Test (maximum score)	+WL Group (n = 18)			−WL Group (n = 17)		
	Mdn	IQR	95% CI	*Mdn*	IQR	95% CI
MLAT (20)	15.50	5.00	[13.50, 17.00]	14.00	4.00	[13.00, 16.00]
LABJ (15)	8.50	8.25	[5.00, 11.00]	8.00	8.50	[5.00, 13.00]
LLAMA (100)	50.00	42.50	[40.00, 80.00]	50.00	45.00	[40.00, 80.00]

Abbreviations: CI, confidence interval; IQR, interquartile range; LABJ, Language Aptitude Battery for Japanese Part 2; *Mdn*. median; MLAT, Modern Language Aptitude Test Part IV; +WL, written languaging; −WL = no written languaging.

Aptitude test results

Descriptive statistics were calculated for the performance of all participants in the two groups on the three measures of aptitude, which are summarized in Table 3. Concerning the MLAT, the medians were 15.50 and 14.00 with interquartile ranges of 5.00 and 4.00 for the +WL group and the −WL group, respectively. None of the participants obtained a perfect score (i.e., 20), 19 being the highest. Meanwhile, one participant scored 6, which was the lowest.

For the LABJ, the medians of the two groups were 8.50 and 8.00, with similar interquartile ranges at 8.25 and 8.50, for the +WL group and −WL group, respectively. Their scores ranged widely from 1 to 15.[5]

The median scores on the LLAMA were 50.00 for both groups with similar interquartile ranges (42.50 and 45.00 for the +WL group and −WL group, respectively). According to the manual (Meara, 2005), scores between 50 and 65 should be interpreted as "good." However, it should be pointed out that the participants' scores varied greatly, ranging from 0 to 100, as reflected in the rather large interquartile ranges.[6]

Turning to inferential statistics, Mann–Whitney tests were run on the scores of the three aptitude tests in order to detect any group differences. None of the tests revealed any statistically significant differences: MLAT, $z = -1.201$, $p = .245$, $d = 0.415$; LABJ, $z = -0.613$, $p = .546$, $d = 0.209$; and LLAMA, $z = -0.368$, $p = .732$, $d = 0.124$.

Correlations between aptitude and L2 learning

To investigate RQ2, a series of Spearman's correlation analyses was performed between the scores on the three aptitude tests and gain scores between the pretest–posttest and pretest–delayed posttest on the two assessments of the +WL group and the −WL group.

As shown in Table 4, statistically significant correlations between the three aptitude test scores and the gain scores of the production and recognition tests were identified only for the −WL group. That is, of all the correlations between the three aptitude test scores and the gain scores of the production and recognition tests, half of them turned out to be statistically significant (6 out of 12; 50%). To be more precise, on the production tests, statistically significant correlations were observed between both measures of inductive language-learning ability (i.e., LABJ, LLAMA) and both the short- and long-term gain scores, with large (short-term gains) and medium (long-term gains) effect sizes. The results indicate that the participants needed a higher inductive language-learning ability to improve their production test scores in both the short and the long term when they had no opportunity to engage in WL. In contrast, the +WL participants achieved similar gains regardless of their aptitude.

As for the recognition tests, statistically significant correlations were found between the long-term gain score and grammatical sensitivity (MLAT) and inductive language-learning ability (LLAMA), both with a large effect size.

TABLE 4 Correlations between aptitude and test gains

	Production tests		Recognition tests	
Test	**G1**	**G2**	**G1**	**G2**
+WL Group (n = 18)				
MLAT				
r	0.444	−0.127	0.381	0.180
95% CI	[−0.036, 0.782]	[−0.631, 0.417]	[−0.199, 0.774]	[−0.409, 0.689]
p	.065	.615	.119	.475
LABJ				
r	0.465	0.154	0.042	−0.023
95% CI	[−0.012, 0.749]	[−0.312, 0.592]	[−0.467, 0.498]	[−0.540, 0.454]
p	.052	.543	.868	.926
LLAMA				
r	0.367	0.290	0.198	0.212
95% CI	[−0.100, 0.771]	[−0.239, 0.723]	[−0.269, 0.605]	[−0.336, 0.713]
p	.135	.243	.431	.398
−WL Group (n = 17)				
MLAT				
r	0.482	0.469	0.125	0.600*
95% CI	[0.052, 0.805]	[−0.046, 0.817]	[−0.371, 0.627]	[0.128, 0.836]
p	.050	.058	.633	.011
LABJ				
r	0.627**	0.540*	0.477	0.318
95% CI	[0.151, 0.909]	[0.051, 0.858]	[−0.013, 0.865]	[−0.206, 0.827]
p	.007	.025	.053	.214
LLAMA				
r	0.717**	0.486*	0.289	0.659**
95% CI	[0.401, 0.898]	[−0.050, 0.773]	[−0.252, 0.712]	[0.296, 0.864]
p	.001	.048	.260	.004

Abbreviations: CI, confidence interval, G1, pre–posttest gain; G2, pre–delayed posttest gain; LABJ, Language Aptitude Battery Part 2; MLAT, Modern Language Aptitude Test Part IV; +WL = written languaging; −WL = no written languaging.
*$p < .05$. **$p < .01$.

DISCUSSION

This study was conducted to investigate the impact of WL on learning and its possible interaction with aptitude abilities. The findings are discussed for each RQ.

The effect of written languaging on L2 learning (RQ1)

The first RQ examined the extent to which WL facilitates learning. The findings of the current study are consistent with those of previous studies (Ishikawa & Révész, 2020; Ishikawa & Suzuki, 2016), but a contradiction was also observed. This section discusses both of the findings, starting with the former.

As reported, the +WL group outperformed the −WL group on both assessments at both posttest sessions. Furthermore, the statistical analyses of the gain scores of the two groups revealed that the +WL group showed significantly improved scores on the delayed posttest for production and on the posttest for recognition, indicating the positive effects of WL on L2 learning and supporting the findings of previous studies (Ishikawa & Révész, 2020; Ishikawa & Suzuki, 2016). These favorable results for the +WL group may be attributed to the following two theoretical assumptions that underlie the rationale for WL. First, the products of WL might be claimed to have benefitted the +WL participants as a source of analysis and reflection on their linguistic issues (Luria, 1999). For example, one participant explained the use of the article "a" in the first sentence, "A crow (. . .) flew to a tree," as "ippon no ki dakara 'a' ['a,' because it is one tree]" (she did not refer to the first mention of the noun "tree"). Then, explaining "the" in the next sentence "A fox walked to the tree," she wrote, "karasu ga iru ki no koto dakara? [because it refers to the tree where the crow is?]" (she correctly explained the use of "the" as anaphoric mention this time). These written languaging episodes (WLEs) seem to demonstrate how the student built on what she wrote (i.e., product), coming to understand the use of articles with her writing-it-through experience (Swain, 2006). Unfortunately, the 5-minute treatment used in the current study did not offer participants additional opportunities to refer to WL products and did not provide enough opportunities to reflect on them. Judging from the products of WL, however, the participants still seem to have benefitted from the opportunity to think through their writing.

Second, as hypothesized, the opportunity to engage in WL is likely to have directed the participants' attention to the target construction, enhancing their noticing and mediating them to think about what is noticed, thereby positively impacting subsequent L2 development. The +WL participants' WLEs seem to support this hypothesis. For example, regarding the use of "a" in " a piece of ∅ cheese (. . .)," one participant wrote, "'a piece of' wa idiomu dakara [because 'a piece of' is an idiom]." Then, explaining the absence of articles before "cheese," he wrote "mae ni 'a' ga aru kara! [because there is 'a' before (piece)!]." The explanation is brief, and his intention is not very clear. However, judging from the exclamation mark, he might have noticed that "a" in "a piece of" is not a meaningless part of the set phrase, but it refers to the number of subsequent pieces of cheese, an uncountable noun. In addition to noticing, from the superior performance of the +WL group over the −WL group, the experience of WL is likely to have induced the deeper processing of +WL participants, resulting in durable and stronger memory representations. If this speculation is correct, it seems to account for the superior performance of the +WL group over the −WL group, especially on the delayed posttest for production.

As stated, one of the findings is counter to those of previous studies (Ishikawa & Révész, 2020; Ishikawa & Suzuki, 2016). The +WL participants outperformed their −WL counterparts in both the production and recognition tests. In contrast, no significant differences were identified regarding the recognition tests in previous studies. This difference may be attributable to the task that the −WL group carried out, that is, copying. Ishikawa and Suzuki (2016) used a grammar exercise, and a silent languaging activity was employed in Ishikawa and Révész (2020).

As mentioned earlier, copying was selected because it involved the act of writing. However, from a SCT perspective, although copying involves imitation, the latter—not the former—mediates development (Vygotsky, 1986). According to Lantolf and Thorne (2006), what is "central to imitation is understanding the goal and the means through which activity is carried out" (p. 167). Emulation is thought to involve understanding the goal but not the means. Thus, the results may indicate that the −WL participants were not "able to match the goal and the means of achieving some activity (imitation)" (Lantolf, 2006, p. 91). Instead, they might have engaged in the copying task "without understanding the relevance of means (emulation)" (Lantolf, 2006, p. 91) or in mere mechanical repetition (mimicry), where "an understanding of the goal is also absent" (Lantolf, 2006, p. 92).

This speculation may be explained by the notion of learner agency (Storch & Wigglesworth, 2010)—that is, the −WL participants might not have found copying to be a meaningful activity, as each is an agent "who perceives, analyses, rejects or accepts solutions offered, makes decisions and so on" (Swain, 2006, pp. 100–101). If this speculation is correct, the act of writing the participants engaged in was probably not "the conscious act of [authentic] writing" (Luria, 1982, p. 167). Given

this possibility, the opportunity for WL might increase the likelihood that learners think and address the task consciously and intentionally; in this way, WL functions as a learning tool (Swain, 2006).

The relationship between written languaging and aptitude (RQ2)

The second RQ asked the extent to which learners' aptitude moderated the positive impact of WL. Statistically significant correlations were observed only for the −WL group. More specifically, of all the correlations between the scores on the three aptitude tests and the gain scores on the two assessment tests, half of them were statistically significant for the −WL group, with medium to large effect sizes. In contrast, no such correlation was found for the +WL group. Thus, the results indicate that aptitude was an important factor for the −WL participants, whereas the +WL participants improved their scores irrespective of their aptitude abilities. Put differently, WL was a more important factor than aptitude for them. These results align with those of Ishikawa and Révész (2023), who also found more statistically significant correlations for the −WL group than the +WL group.

This outcome seems to be explained by previous ATI research (Li, 2019). The impact of aptitude on learning has been identified as different depending on the nature of instruction and treatment (i.e., the degree of implicitness or explicitness). Therefore, the experience of WL—a combination of writing and languaging, both conscious processes (Luria, 1999; Swain, 2006)—is likely to have made the nature of the treatment more explicit for the +WL participants than for the −WL participants, creating an optimal condition for explicit learning. In addition, considering that the slower pace of writing may provide learners with access to explicit knowledge (Williams, 2008), the +WL participants might have used explicit knowledge as a tool (Ellis, 2004) when they addressed the treatment task. These factors were likely to have leveled out any effects for aptitude, resulting in no significant correlations between aptitude and test gains. In contrast, in terms of the −WL group, who had no access to WL, only the participants with higher aptitude were likely to have achieved gains from the treatment through the use of explicit knowledge. If this interpretation is correct, it seems to account for the statistically significant correlations between the aptitude test scores and the gain scores observed only for the −WL group.

It is worth mentioning that this line of reasoning seems to be in line with the claim of Cronbach (1967), who suggested that erasing individual differences is one way to adapt instruction to individual differences. Given the findings of the present study (i.e., no statistically significant correlations between aptitude and gain scores for the +WL group, whereas half of the correlations turned out to be significant for the −WL group), WL appears to have succeeded in erasing individual differences, mediating the mechanism by which the +WL participants learn L2 grammar. As stated, aptitude has been considered to be a fixed and immutable trait (Carroll, 1981). Contradicting the mainstream view, however, the findings of the current study seem to indicate that, in fact, aptitude is modifiable and depends very much on the interaction between the person and the environment of the activity that the person is engaged in, such as WL. Simply put, WL might have changed the participants' aptitude, as evidenced by the effects of WL. If this speculation is correct, aptitude may need to be redefined as a construct that is sensitive to sociocultural contexts.

Meanwhile, as stated earlier, the difficulty of a target construction has been suggested as another factor that influences the impact of aptitude (Li, 2013; Skehan, 2015). Ishikawa and Révész (2023) observed results similar to those of the current study with the past counterfactual conditional as the target construction. Comparatively, the target in this study (i.e., articles) is semantically complex but syntactically simpler. In addition, this study focused on only one use of the "a"–"the" distinction, of which the participants were assumed to have prior knowledge. Thus, although this distinction is considered problematic because of L1–L2 differences and its nonsaliency, it was likely to be within the +WL participants' processing capacity. Accordingly, as Li (2013) predicted, participants were likely to have figured out the grammar rule of articles when they examined the model text with only the aid of WL, regardless of their aptitude abilities.

This reasoning seems to be consistent with the condition of Stefanou and Révész (2015), one of the two studies that examined correlations between the impact of corrective feedback and aptitude abilities with English articles as the target construction. This study identified more significant correlations between learning from implicit feedback and aptitude, and the participants were EFL high school students with some prior knowledge of the target feature. In contrast, Sheen (2007) observed more significant correlations between L2 learning from explicit feedback and aptitude with ESL learners lacking prior knowledge of the same target. Given these, of the three factors (i.e., high aptitude, explicit instruction, prior knowledge), learners are likely to need at least two to develop proficiency. Namely, when learners have some prior knowledge and receive explicit instruction, as in Stefanou and Révész, they do not seem to need a high aptitude to develop proficiency. In contrast, when learners have no prior knowledge, as in Sheen's study, they are likely to need both explicit instruction and a high aptitude. This speculation aligns with the results of this study. Although all the participants had some prior knowledge, the +WL participants did not seem to need a high aptitude to develop proficiency, most likely because of the explicit nature of WL. Meanwhile, the −WL participants, who were not given the opportunity of WL, are likely to have needed a high aptitude to develop proficiency. Thus, in addition to target difficulty, learners' prior knowledge seems to be another critical factor to be considered.

Limitations and future directions

Before concluding, several limitations of this study should be acknowledged for future research. First, this study started out with 73 participants, but the final number was 35. A larger pool of participants would have provided more reliable data. Second, for practical reasons, this study did not provide enough opportunities for learners to engage in reflection and did not examine the effects of WL to the full extent possible. Given that WL is expected to benefit learners through its process and product, future research should provide more ample opportunities to utilize the product of their WL. Third, some may wonder about the lack of a control group and whether the observed results might have been due to test repetition effects. While this could be seen as a serious shortcoming of the study, we point out that the +WL group still outperformed the −WL group despite the same repetition. Thus, we argue that test repetition did not greatly impact the results. Fourth, the delayed posttest was administered only 1 week after the treatment, not measuring the long-term effect in a strict sense. Fifth, the two assessments employed in the current study were highly controlled. Thus, the participants are likely to have addressed the tests, drawing on explicit knowledge alone. Given that assessment tasks seem to be an influential factor (Li et al., 2019), future research should consider task effects. Namely, the possible link between the impact of WL and aptitude should be examined further, including types of tasks that are likely to tap into implicit knowledge. Sixth, as in Ishikawa and Révész (2023), the target construction may have been within the learners' processing capacity as only one aspect was tested in this study. Future research should investigate the correlation between the effects of WL and aptitude with more target constructions, especially those that are likely to be outside their processing capacity due to L1–L2 differences, syntactic or semantic complexity, or unfamiliarity to the learners.

CONCLUSION

This study investigated the effects of WL on L2 learning and the extent to which learners' aptitude abilities mediate such effects. Two major findings emerged. First, WL was identified as benefitting learners, which is in line with previous research. The findings support Swain's (2006) statement that languaging in the written modality facilitates learning. Although writing and WL are not identical, given the results, the +WL participants might have utilized WL as "a powerful instrument of thought" (Luria, 1999, p. 103), benefitting from the process and product of WL. Second, the explicit nature of WL has likely neutralized learners' aptitude, at least to some degree, benefitting the +WL participants

regardless of their aptitude. The findings suggest that WL changed their aptitude, indicating that aptitude is modifiable and depends on the social context.

The following pedagogical implications can be drawn from the findings of the current study. First, WL is likely to be a beneficial activity for learners as it potentially changes their aptitude. Thus, it seems to be a valuable instructional and learning tool for any learner and any setting given its high practicality (as no special equipment is required). Second, the product of WL benefits not only learners but also teachers as a rich source of information regarding their students' development. Namely, students' WL enables teachers to provide (a) timely and appropriate feedback on the efficacy of their teaching, and (b) a chance to reflect on and adjust their teaching promptly when necessary.

Research on the relationships between the impact of WL on learning and learners' differences (e.g., aptitude) has just begun. Given the potential significance of individual differences in instruction, investigating the link between the two (i.e., WL and aptitude) seems to be a promising avenue of research. More studies should be conducted with various task types, target constructions, and learners with various profiles. In addition, more research should be conducted to identify the process in which written languaging and learner factors such as aptitude may dynamically interact and impact each other. We hope that this study will serve as a stepping stone for future research.

ACKNOWLEDGMENTS

We would like to thank the editors of this special issue, James Lantolf and Matthew Poehner, as well as anonymous reviewers for their thoughtful comments and helpful suggestions on earlier drafts of this paper. Our thanks also go to the participants of this project. This work was supported by JSPS KAKENHI Grant Numbers 17K02979 and 20K00865.

ORCID

Masako Ishikawa https://orcid.org/0000-0002-7035-0618

ENDNOTES

[1] Through languaging, learners solve their affective and cognitive issues, which is beyond the scope of this paper (see Swain, 2010).

[2] It was decided to allocate 5 minutes for WL and copying based on the pilot conducted with learners similar to the current study's participants in terms of proficiency and background.

[3] Both the +WL and −WL participants could have engaged in a form of private speech. However, this discussion is beyond the scope of this study, as audio recording data was not collected.

[4] From an SCT perspective, imitation plays a major role in development (Vygotsky, 1986). Thus, it could be argued that engaging in the copying task positively impacted L2 learning and development (please refer to "Discussion" section).

[5] Although medians were not reported in the study by Shintani and Ellis (2015), who also employed LABJ Part 2, the means ranged from 13.57 to 14.23 out of 15 across four treatment groups, with rather small standard deviations ranging from 0.82 to 1.44.

[6] Grañena (2013) reported that the average LLAMA score was 56.67 ($SD = 24.09$).

REFERENCES

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

Carroll, J. B., & Sapon, S. M. (1959). *Modern Language Aptitude Test*. The Psychological Corporation.

CelceMurcia, M., & Larsen–Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course*. Heinle & Heinle Publishers.

Cronbach, L. J. (1967). How can instruction be adapted to individual differences? In R. M. Gagne (Ed.), *Learning and individual differences* (pp. 23–39). Merrill Books.

Cummings, M. C. (2004). "Because we are shy and fear mistaking:" Computer mediated communication with EFL writers. *Journal of Basic Writing*, *23*, 23–48. https://www.jstor.org/stable/43444066

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 612–630). Blackwell.

Ehrman, M. E., & Oxford, R. L. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal*, *79*, 67–89. https://doi.org/10.1111/j.1540-4781.1995.tb05417.x

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*, 227–275. https://doi.org/10.1111/j.1467-9922.2004.00255.x

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal*, *63*, 97–107. https://doi.org/10.1093/elt/ccn023

Gal'perin, P. I. (1989). Mental actions as a basis for the formation of thoughts and images. *Soviet Psychology*, *27*, 45–64. https://doi.org/10.2753/RPO1061-0405270345

Gal'perin, P. I. (1992). Stage-by-stage formation as a method of psychological investigation. *Journal of Russian & East European Psychology*, *30*, 60–80. https://doi.org/10.2753/RPO1061-0405300460

Grañena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. In G. Grañena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105–129). John Benjamins.

Ishikawa, M., & Révész, A. (2020). L2 learning and the frequency and quality of written languaging. In W. Suzuki & N. Storch (Eds.), *Languaging in language learning and teaching: A collection of empirical studies* (pp. 220–240). John Benjamins.

Ishikawa, M., & Révész, A. (2023). Written languaging, learners' aptitude, and L2 learning through dictogloss tasks. In S. Li (Ed.), *Individual differences in task-based language learning and teaching*. John Benjamins.

Ishikawa, M., & Suzuki, W. (2016). The effect of written languaging on learning the hypothetical conditional in English. *System*, *58*, 97–111. https://doi.org/10.1016/j.system.2016.02.008

Lantolf, J. P. (2006). Sociocultural theory and L2: State of the art. *Studies in Second Language Acquisition*, *28*, 67–109. https://doi.org/10.1017/S0272263106060037

Lantolf, J. P., & Poehner, M. E. (2023). Sociocultural theory and classroom second language learning in the East Asian Context: Introduction to the special issue. *Modern Language Journal*, *107*,(Supplement 2023), 3–23. https://doi.org/10.1111/modl.12816

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.

Leow, R. P. (2015). *Explicit learning in the classroom: A student-centered approach*. Routledge.

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

Li, S. (2019). Six decades of language aptitude research: A comprehensive and critical review. In Z. Wen, P. Skehan, A. Biedron, S. Li, & R. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research, and practice* (pp. 78–96). Routledge.

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics*, *40*, 693–722. https://doi.org/10.1017/S0142716418000796

Luria, A. R. (1982). Basic forms of the speech utterance: Oral (monologic and dialogic) and written speech. In J. Wertsch (Ed.), *Language and cognition* (pp. 159–168). Wiley.

Luria, A. R. (1999). Speech development and the formation of mental processes. In P. Lloyd & C. Fernyhough (Eds.), *Lev Vygotsky: Critical assessments, volume II. Thought and language* (pp. 84–122). Taylor & Francis.

Manchón, R. M. (2011). Writing to learn the language: Issues in theory and research. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 61–82). John Benjamins.

Master, P. (1997). The English article system: Acquisition, function, and pedagogy. *System*, *25*, 215–232. https://doi.org/10.1016/S0346-251X(97)00010-9

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

Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning*, *50*, 617–673. https://doi.org/10.1111/0023-8333.00142

Pica, T. (1983). Methods of morpheme quantification: Their effect on the interpretation of second language data. *Studies in Second Language Acquisition*, *6*, 69–78. https://doi.org/10.1017/S0272263100000309

Pimsleur, P. (1966). *The Pimsleur Language Aptitude Battery*. Harcourt, Brace, Jovanovic.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. https://doi.org/10.1111/lang.12079

Reichelt, M. (2009). A critical evaluation of writing teaching programmes in different foreign language settings. In R. M. Manchón (Ed.), *Writing in foreign language contexts* (pp. 183–206). Multilingual Matters.

Robinson, P. (2002). Learning conditions, aptitude complexes, and SLA: A framework for research and pedagogy. In P. Robinson (Ed.), *Individual differences in instructed language learning* (pp. 113–133). John Benjamins.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 45–73. https://doi.org/10.1017/S0267190505000036

Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence: Quantitative and qualitative analyses*. Peter Lang.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, *41*, 255–281. https://doi.org/10.1002/j.1545-7249.2007.tb00059.x

Shintani, N., & Ellis, R. (2015). Does language analytical ability mediate the effect of written feedback on grammatical accuracy in second language writing? *System*, *49*, 110–119. https://doi.org/10.1016/j.system.2015.01.006

Skehan, P. (2015). Foreign language aptitude and its relationship with grammar: A critical overview. *Applied Linguistics*, *36*, 367–384. https://doi.org/10.1093/applin/amu072

Stefanou, C., & Révész, A. (2015). Direct written corrective feedback, learner differences, and the acquisition of second language article use for generic and specific plural reference. *Modern Language Journal*, *99*, 263–282. https://doi.org/10.1111/modl.12212

Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake and retention of corrective feedback on writing. *Studies in Second Language Acquisition*, *32*, 303–334. https://doi.org/10.1017/S0272263109990532

Suzuki, W. (2009). Improving Japanese university students' second language writing accuracy: Effects of languaging. *Annual Review of English Language Education in Japan*, *20*, 81–90. https://doi.org/10.20581/arele.20.0_81

Suzuki, W. (2012). Written languaging, direct correction, and second language writing revision. *Language Learning*, *62*, 1110–1133. https://doi.org/10.1111/j.1467-9922.2012.00720.x

Suzuki, W. (2016). The effects of quality of written languaging on second language learning. *Writing & Pedagogy*, *8*, 461–482. https://doi.org/10.1558/wap.27291

Suzuki, W., & Storch, N. (Eds.). (2020). *Languaging in second language learning and teaching*. John Benjamins.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–81). Cambridge University Press.

Swain, M. (2006). Languaging, agency and collaboration in advanced second language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95–108). Continuum.

Swain, M. (2010). Talking it through: Languaging as a source of learning. In R. Batstone (Ed.), *Sociocognitive perspectives on language use and language learning* (pp. 112–130). Oxford University Press.

Swain, M., Lapkin, S., Knouzi, I., Suzuki, W., & Brooks, L. (2009). Languaging: University students learn the grammatical concepts in French. *Modern Language Journal*, *93*, 5–29.

Terauchi, H. (2017). English education at universities in Japan: An overview and some current trends. In E. S. Park & B. Spolsky (Eds.), *English education at the tertiary level in Asia* (pp. 65–83). Routledge.

Vygotsky, L. S. (1986). *Thought and language*. MIT Press.

Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber, A. S. Carton, & N. Minick (Eds.), *The collected works of L. S. Vygotsky. Volume 1: Problems of general psychology* (pp. 37–285). Plenum Press.

Williams, J. (2008). The speaking–writing connection in second language and academic literacy development. In D. Belcher & A. Hirvela (Eds.), *The oral/literate connection: Perspectives on L2 speaking, writing, and other media interactions* (pp. 10–25). University of Michigan Press.

Williams, J. (2012). The role(s) of writing and writing instruction in L2 development. *Journal of Second Language Writing*, *21*, 321–331.

Yalçın, Ş., & Spada, N. (2016). Language aptitude and grammatical difficulty: An EFL classroom-based study. *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

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Masako, I., Wataru, S. (2023). Effects of written languaging on second language learning: Mediating roles of aptitude. *Modern Language Journal*, *107,* 95–112. https://doi.org/10.1111/modl.12817